N° d'ordre : | 1 | 1 | 5 |

ÉCOLE CENTRALE DE LILLE

THÈSE

présentée en vue d'obtenir le grade de

DOCTEUR

en

Spécialité : Automatique et Informatique Industrielle

par

Hanal ABU ZANAT

Doctorat délivré par l'École Centrale de Lille

Modèles d'équité pour l'amélioration de la qualité de service dans les réseaux sans fil en mode ad-hoc

Fairness models to improve the quality of service in ad-hoc wireless networks

Soutenance le 10 DEC 2009 devant le jury d'examen :

Président Rapporteur	Hossam AFIFI	Professeur à la Télécom SudParis
Rapporteur	Philippe LE PARC	Professeur à l'Université de Bretagne Occidentale
Examinateur	Christophe CHASSOT	Professeur à l'INSA de Toulouse
Examinateur	Jean VAREILLE	Maître de Conférences à l'Université de Bretagne Occidentale
Examinateur	Mohamed GHAZEL	Chargé de recherche à l'INRETS Villeneuve d'Ascq
Directeur de thèse	Armand TOGUYENI	Professeur à l'École Centrale de Lille
Co-directeur de thèse	Benoit TROUILLET	Maître de Conférences à l'École Centrale de Lille

Thèse préparée au sein du Laboratoire LAGIS

École Doctorale SPI 072

À mon père

À ma mère

À mon frère Raed

À mes sœurs Manal, Maram et Hadeel

Remerciements

En premier lieu, je tiens à exprimer ma profonde gratitude aux directeurs de cette thèse le Professeur Armand TOGUYENI et à Benoit TROUILLET, Maître de Conférences. Je veux les remercier pour leur encadrement plein d'enthousiasme et de rigueur et pour la confiance dont ils ont fait preuve à mon égard. De plus, je tiens à leur manifester ma sincère reconnaissance pour leurs grandes qualités humaines qui ont fait que cette thèse se passe dans la bonne humeur.

Finalement, ces remerciements ne seraient pas complets si je n'y associais pas toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce travail, en particulier, mes collègues, tout le personnel du LAGIS et de l'École Centrale de Lille pour leur bonne humeur et leur disponibilité, notament: Virginie Leclercq et Régine Duplouich.

Contents

List of Figures

List of Tables

Fairness models to improve the quality of service in ad-hoc wireless networks

H. A. ABUZANAT

December 2009

General Introduction

The Internet, that connects all users, has many topologies, infrastructures and protocols for wired and wireless networks. However, a big challenge is faced to develop and implement their performance and their characteristics, which can permit all users to transfer their packets of different applications: data packets (ex. FTP), control packets (ex. Admission control), and media packets (ex. Voice 'VoIP', video streams), real-time packets (ex. safety-critical packets), etc.

One of the most famous protocols that became a dominant protocol in the internet world is IP (Internet Protocol) (RFC7 81a). IP was designed to provide a Best-effort service, which is the service in the internet without priority or guarantees, for delivery of data packets and to run across any network transmission media and system platform. These things are increasing the popularity of IP and make it carrying everything over IP; ex. Real-time data, audio, streaming video, etc. So, this makes IP to transfer all packets of different applications.

However, at each moment in the Wide Area Network (WAN), the network becomes a heavy loaded. That means, WAN bandwidth is not free yet. As a consequence, the network traffic is not all equally and deployed fairly with respecting the Service Level Agreement (SLA). Therefore, time-sensitive applications (video, audio, real-time streams) need a guarantee for transmission with end-to-end bounded delay.

In addition, IP is not the best protocol that control every thing and managing the entire network. Thus, the misbehaved (mischievous, ill-behaved or annoying behavior) users can consume the bandwidth resources more than the other users. For example, a user is sending FTP packets of size 1500byte and another user is sending Real-time packets of size 512byte (normally, the packet size of real-time is less than 512byte). So, the ill-behaved users lead to a fairless network, because the router will serve the packets in round-robin manner. That means the packets of FTP will consume almost 3 times of bandwidth resources more than the real-time packets. Further, the heavy traffics that were caused by the ill-behaved users, make the congestion and bottle-neck in the network. This leads to a starvation. So, if the traffic is regulated with fairness, it could resolve some of crucial problems in the networks like the congestion and starvation.

Indeed, the heavy traffic, congestion or starvation are undesirable for feedback-based congestion control algorithms in data communication networks. To solve this problem, the senders have to balance between two considerations. On the one hand, when they want to send data to the network as fast as possible. On the other hand, when they do not want to send the data so fast. To achieve the best performance, the source needs to detect the amount of bandwidth available to itself and match its sending rate to the available bandwidth.

A traffic contract as Service Level Agreement (SLA) specifies guarantees for the ability of a network and its protocols to give guaranteed performance, throughput and bounded delays. This could be made by prioritizing traffic. Therefore, the network requires Quality of Service (QoS) (Shin 03; Ciscc; RFC1; RFC2a; Cisc 05) in addition to best-effort service. QoS has the permission to control and manage the entire network. Further, different applications have varying needs for delay, delay variation (jitter), bandwidth, packet loss, and availability. These parameters form the base of QoS. Thus, the IP networks should be designed to provide the request of the applications for QoS. Therefore, the networks must provide secure, predictable, measurable and guaranteed services to these applications. So, the required service can be achieved by managing the delay, delay variation (jitter), bandwidth and packet loss parameters on a network, while maintaining its simplicity and its scalability. Which is critical and not simple to achieve, specially for the entire network complex applications.

A defined Quality of Service may be required for certain types of network traffic. For example, IP telephony (VoIP) may require strict limits on jitter and delay. And also, a streaming multimedia may require guaranteed throughput. Further, some types of service applications are called inelastic, meaning that they require a certain level of bandwidth to function, any more than required is unused, and any less will make the service non-functioning. For example a safety-critical application, such as remote surgery, which is the ability for a doctor to perform surgery on a patient even though they are not physically in the same location. That must require a guaranteed service level of availability with no permission of failure or malfunction. Also, this is called hard QoS. By contrast, elastic applications can take advantage of however much or little bandwidth is available.

So, the problem is how to guarantee their QoS, which poses serious challenges to network researchers, whether from the service providers or internet end customers. For example, (Zhon 07; Parc 05; Vare 03) pose the problem of real-time control of machinery on the internet. Basically, this requires QoS guarantees bounded delays to have a deterministic time.

Therefore, one of the topics that can enhance the QoS guarantees is the fairness. As described, regulating traffic with fairness is one of the missions of QoS in all types of networks. Thus, QoS need to allocate the resources with fairness depending on the priority of packets to guarantee the services for time-sensitive applications. However, the fairness objective in wireless network depends on its topology, its size, its technology and the type of traffics and their applications. Therefore, it must follow an agreement (i.e. SLA). However, the fairness objective could be referred to many considerations; such as:

- Flow fairness: the throughput of each flow must be allocated fairly between all flows (No priority).
- Prioritized flow fairness: the throughput of each flow must be allocated fairly between all flows of the same priority level.
- Source fairness: the throughput of each source node must be allocated fairly between all the sources.
- Reserved source fairness: the bandwidth of each source node must be allocated fairly between all the nodes. The difference between this definition and the last one is that not all the nodes are source nodes at the same time.
- Max-Min fairness: it allocates low throughputs required for some sources. And the rest of bandwidth is distributed for the other sources, which require high throughputs, and so on.

Further, nowadays, the technologies of connecting people is enhanced. So, you would be connected wherever you are, and whatever you want to send (i.e. real-time, video, audio, etc.). And also, these technologies could be implemented on any infrastructure networks. Therefore, in the wired network side, there is a large research on optimizing QoS. That enhancement of QoS could be considered not bad and better than the enhancement of QoS in wireless networks side. Although there is a lot of research concerning the QoS optimization in wireless networks, this optimization does not have at least the same level compared with that optimization in wired networks. This is referred to the problematic of wireless networks is not the same as that in wired networks. So, we will describe in brief the problematic of wireless networks. For example, the channel error in wireless networks is happened often when there is an interference with other channels or no Line of Sight (LoS) because of the obstacles that attenuate the wireless signal. Also, for the same mentioned reasons and the mobility, the wireless networks bandwidth is variable and makes the required throughputs hard to be allocated fairly.

Further, some of wireless devices have limited energy (i.e wireless sensors, mobiles, PDAs, etc.) that makes the routing paths changeable with extra control traffic.

Moreover, the wireless networks problematic leads also to face many difficulties for adopting the enhancement of wired side to wireless side. In addition, the definitions of OSI layers (MAC and PHY layers) are translated and adopted in the wireless networks to still have the compatibility, because the functionality of accessing channel in wireless network depends on these layers. Therefore, new phenomenas are appeared (i.e. hidden and exposed terminal problems, spatial reuse, unfairness, etc.) and need solutions. Indeed, an extension of MAC layer and dynamic solutions are needed to resolve almost the problematic of wireless networks. Unfortunately, like these solutions will have complex calculations and heavy traffic. Therefore, there is a trade-off between the perfect solution and the complexity. However, there is no perfect solutions that resolve the wireless network problematic, because every solution has its undesired effects. So, it is better to resolve each problem separately with its circumstances to achieve the required goals and evaluate its benefits. Therefore, the wireless networks can operate in two modes. One of them is the infrastructure mode targeted to the centralized purpose. And this type of network can be managed centrally. Thereby, the most of research of this mode is applied on the center devices. The other one is ad-hoc mode targeted to the distributed purpose. And this type of network could be managed independently by each device.

The main goal of wireless ad-hoc networks is to allow a group of communication devices to setup and maintain a network among themselves, without the support of a central controller. Thereby, the collaboration between these devices establish the connectivity between them. However, some of situations (i.e. home and office networks, conference meetings, military applications, crisis response for emergency, etc.) (Jurd 07) that require quick local network deployment without infrastructure, and guaranteed service for receiving their data.

Therefore, the most of the previous undesired effects are appeared in this type of wireless networks. So, if some of these undesired effects are treated, a real enhancement will be achieved. This enhancement will make serious problems to be resolved. One of these serious problems is satisfying and enhancing the QoS in wireless networks especially ad-hoc mode.

So, in our research, we aim to enhance and optimize the QoS in wireless ad-hoc network with fairness. Thus, multiple models are proposed to enhance the QoS in ad-hoc. This enhancement is proved by implementing our models and evaluating them. However, our models are evaluated and compared with recent standards of QoS in wireless networks to obtain the real benefits of our enhancement.

In our thesis, we will first introduce the wireless network technologies. Then we present a state of the art of the QoS, its characteristics, its mechanisms, its protocols and its researches, in both wired and wireless environments. Then, the second part is targeted to our models that enhance QoS with fairness (i.e. F-EDCA (Abuz 09c; Abuz 08) and FQ-EDCA (Abuz 09b; Abuz 09a)). These models are implemented and simulated by using the Network simulator (NS-2). As a consequence, an evaluation is obtained for each model that present the concerned enhancement. Finally, we will compare these two models and present our conclusions.

Part I

The State of The Art

Chapter 1

Wireless Networks

1.1 Introduction to Wireless Networks

The wireless networks appeared in the last decades. It refers to the interconnection between the nodes (hosts) without wires, using radio waves. These waves are propagated with an infrared band or an unlicensed spectrum bands frequencies (2.4GHz and 5GHz). The frequency of 2.4GHz is associated globally for Industrial, Scientific and Medical purpose (ISM band), while the frequency of 5GHz is associated for Unlicensed National Information Infrastructure (UNII band). These bands frequencies are unlicensed and free, which make the wireless networks widely deployed. However, this evolution of networks from wires to wireless contributes to have many advantages such as:

- Facilitating the construction of networks and flexibility.
- Minimizing the budget of wiring and installations.
- Interconnecting the mobile objects.
- Being used for variable topologies.

On the other hands, many disadvantages need to be overcome, such as:

- Low bandwidth compared to the wired networks.

- Hard challenge to adapt the wired protocols.

So, theses disadvantages are problematic in wireless networks, and caused by many reasons such as the variation of bandwidth is depending on the network's topology, the mobility and the density of the wireless devices, the channel-error because of the signal attenuation (i.e. between the buildings, trees, mountains, etc.). As a consequence, these problems make the optimization of the wireless networks harder and more difficult.

Therefore, various technologies of wireless networks are developed and associated for computer networks and its devices. Generally, these technologies are classified depending on the size of networks or type of using (i.e. WPAN, WLAN and WMAN) and standardized in IEEE802.1xx. Where, xx is different numbers and characters that specifies various series of wireless technologies, see table 1.1 and table 1.2. One of them is the standard IEEE 802.11. That is the most popular standard in wireless computer networks. However, the difference between these technologies is the data rate, signal frequency, type of carrier, quality of service, etc. They have different characteristics implemented and taken place in the physical (PHY) and the data-link (MAC) layers. Where, the definition of the network layers is based on the OSI model.

1.2 Wireless Network Technologies

We will describe the technologies of wireless networks depending on their coverage area.

1.2.1 WPAN

WPAN (Labi 06) is the Wireless Personal Area Networks. It refers to small sized networks that are used to establish connections between a computer and its devices which covers an area of only a few dozen of meters (i.e. room).

This network is standardized by IEEE802.15. We will describe the most important technologies of WPANs.

Bluetooth

Bluetooth (Labi 06) is a vast commercial name specified by the standard IEEE802.15.1 (IEEE 05b). This technology is broadly used to interconnect the devices of computers (i.e. headphone, mouse, keyboard, mobile phones, etc.) used clearly for personal purposes. This technology is operated in the ISM band. Indeed, it uses a frequency hopping spread spectrum signaling method (FHSS), while the original standard of IEEE802.11 uses two variations FHSS and DSSS. The last one is direct sequence spread spectrum signaling method. Recently, the new technology of WLAN or Wi-Fi devices uses DSSS. That means, the Bluetooth does not interfere with the new Wi-Fi devices of WLAN, although they use the same band.

ZigBee

ZigBee (Labi 06) is targeted for WPAN applications that require a low data rate, long battery life, and secure networking. Theses characteristics make it to have low prices. This technology has a large deployment and it is adapted for sensor networks. Generally, it is standardized with IEEE802.15.4 (IEEE 06). It uses the ISM band like WLAN or Wi-Fi devices. And also, it uses the DSSS signaling method.

1.2.2 WLAN

WLAN (Labi 06) is the Wireless Local Area Networks. It is the most famous wireless networks. That represents the basic standard of IEEE802.11 (IEEE 97). It is referred to the commercial term (Wi-Fi). However, Wi-Fi is a trademark registered for WECA (Wireless Ethernet Compatibility Alliance), with which the wireless devices of WLAN based on IEEE802.11 are certified. This certification grants the interoperability between the different wireless devices of WLAN.

There are several series of the IEEE802.11 family. So, we will describe the most popular and concerned ones:

- IEEE802.11: it is the legacy standard associated to WLAN. The data is carried over frequency hopping spread spectrum (FHSS) or direct sequence spread spectrum (DSSS) signal modulation. It can provide a data rate of 1Mbps or 2Mbps in the ISM band (2.4GHz).

- IEEE802.11a: it is an extension of IEEE802.11. It provides a data rate, depending on the channel conditions, that could be any of 6, 9, 12, 18, 24, 36, 48, or 54 Mbps. Whereby, another signal modulation is adopted rather than FHSS or DSSS, which is the orthogonal frequency division multiplexing (OFDM). Also, it is applied in the UNII band (5GHz) that supports 8 to 12 non-overlapping channels.

- IEEE 802.11b: it is an extension of IEEE802.11. Also, it is considered as IEEE802.11 High Rate. It is referred to the commercial term (Wi-Fi). It provides a data rate, depending on the channel conditions, that could be any of 1, 2, 5.5, or 11 Mbps. It uses only DSSS signal modulation scheme in the ISM band (2.4GHz) that supports 3 non-overlapping channels.

- IEEE 802.11g: it is an extension of IEEE802.11 (Wi-Fi). It uses the OFDM, DSSS or both signal modulations to obtain on the data rate in range of 5.5 to 54Mbps in the ISM band (2.4GHz).

The versions of (a, b and g), as mentioned above, describe the physical layer specifications. Thus, these versions have different specifications though they use a common MAC layer. Further, the difference of the Inter-Frame spaces between them is considered, as seen in table 1.3. And they have different signal modulations, as seen in table 1.2.

Table 1.1: The characteristics of the WLAN standards

The standard	Release date	Maximum BW (Mbps)	Frequency Band (GHz)
802.11	1997	2	2.4
802.11a	1999	54	5
802.11b	1999	11	2.4
802.11g	2003	54	2.4
802.11e	2005	11;54	2.4;5
802.11n	2009	540	2.4;5

- IEEE 802.11e: it defines the Quality of Service (QoS) supported for WLANs. It is an enhancement to the legacy IEEE802.11a/b/g. That means, it uses the PHY layer specifications of IEEE802.11a/b/g and enhances the MAC layer. This enhancement can support the QoS guarantees for the real-time and the multimedia applications and maintain the compatibility with IEEE802.11a/b/g standards for WLAN. It will be described in details in the state of the art (see section 2.4).

- IEEE 802.11n: it is a work in progress that supports data rates over than 100 Mbps by adding multiple-input multiple-output (MIMO) antennas instead of one. The additional transmitter and receiver antennas allow for increased data throughputs. This version of IEEE802.11 is considered as an enhancement for the legacy standard of IEEE802.11a/g that uses the signal modulation OFDM. And secondly by adding the channel-bonding operation applied in the physical (PHY) layer, which uses two non-overlapping channels to increase data transmission. And, a frame aggregation is applied to the MAC layer. So, by these methods the bandwidth obtained is more than the legacy IEEE802.11a/g multiple times and reach to 540Mbps.

However, in this thesis, we focus our studies in enhancing generally the standard IEEE802.11e specially in ad-hoc mode. And also, the other technologies of IEEE802.11 (i.e. a,b,g, etc.) almost have an identical behaviors when there is a congestion, heavy load or bottleneck. So, all the simulations will be performed with the characteristics of the standard IEEE802.11b.

1.2.3 WMAN: WiMAX

WMAN (Labi 06) is the Wireless Metropolitan Area Networks. It refers to a large sized network, which covers an area of dozens of kilometers. This type of network, as a consequence of the explosion and deployment of the wireless networks, promises a broadband wireless networks comparable with the wired DSL or cable networks. One of the most new technologies of WMAN is WiMAX standardized by IEEE802.16 (IEEE 05c).

WiMAX is the Worldwide Interoperability for Microwave Access. Commercially, it is deployed and standardized with a subset of the IEEE802.16 standard. This standard provides a broadband wireless access for fixed station and named fixed WiMAX. While, an amendment of the IEEE802.16 standard is achieved to support the mobile users and named the mobile WiMAX, which is a subset of the IEEE802.16e standard. However, the mobile users can be connected to this broadband wireless access by extending the specifications of WiMAX and supporting the Wi-Fi hot-spots with internet. In addition, it adopts several technologies of high performance physical layer (PHY). This is the major technical advance of WiMAX to provide a high data rate connections (i.e. tens of Mbps) between the base station (BS) and the subscriber stations (SS) for distance of several kilometers.

Also, WiMAX adopts OFDMA (Orthogonal Frequency Division Multiple Access) as signal modulation that is based on OFDM. Further, mobile WiMAX uses a scalable OFDMA (SOFDMA) for broadband wireless access to mobile users.

Another enhancement in PHY layer, WiMAX implements multiple antennas for the transmitter and the receiver (MIMO) further SOFDMA. Thus, it can support of data rate over than 100Mbps. Therefore, it can cover thousands of users and mobile users distributed in a wide area.

Table 1.2: The comparison between the standards of WPAN, WLAN and WMAN

Technology	WMAN	WLAN	WPAN	
Commercial name	WiMAX	Wi-Fi	Bluetooth	ZigBee
Standards	IEEE802.16e	IEEE802.11 (a,b,g,n)	802.15.1	802.15.4
Release year	2005	1999, 1999, 2003, 2009	2002	2006
Frequency Band	2-66GHz	2.4GHz, 5GHz	2.4Ghz	2.4GHz
Coverage	Wide area 50km	Local area 100m-2km	Personal area 10m	
Maximum BW	>100Mbps	540Mbps	3Mbps	250kbps
Number of users	Thousands	Dozens	<Dozen	Dozens
Transmission Method	MIMO, OFDMA	MIMO, OFDM, DSSS, FHSS	FHSS	DSSS

A comparison that summarize the difference between the various technologies of the wireless networks, is presented in table 1.2.

1.3 The operation modes in WLAN

IEEE802.11 can operate in two modes:

1.3.1 The Infrastructure Mode

The infrastructure or centralized mode is a centralized wireless network that has the access points (AP) or base stations (BS). The AP is responsible to associate and administrate the wireless devices. These devices have many synonyms, such as stations (STA), remote stations (RS), mobile stations (MS), terminal hosts or nodes. So, the communications between two STAs must be coordinated through the AP. And also, the AP can interconnect the interfaces between the STAs with the wired networks for LAN and WAN (i.e. Router AP), as seen in figure 1.1.b.

1.3.2 Ad-hoc Mode

Ad-hoc or distributed mode is a distributed wireless network. An ad-hoc network is an autonomous wireless network that can be formed without the need of any infrastructure, centralized administration or centralized coordinator. This means, it does not need an AP. It is composed of stations (STAs) or nodes. Every node communicates with each other in a peer-to-peer fashion through single-hop or multi-hop paths. Therefore, a node can route and forward the packets of others nodes. They can cooperate among them to achieve the best connection between the sender and receiver, see figure 1.1.a.

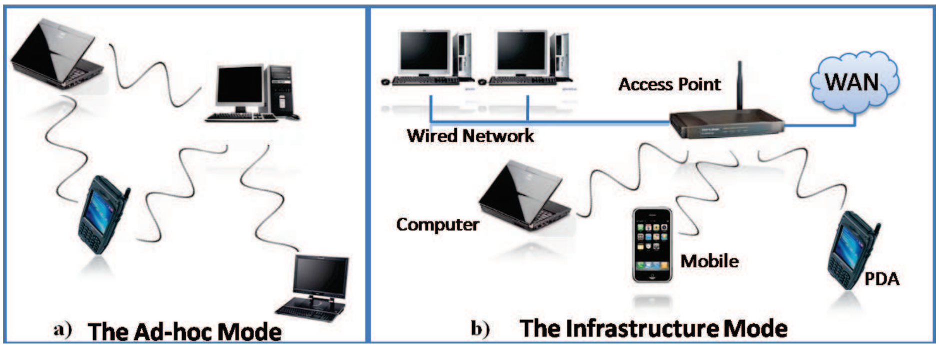


Figure 1.1: The operation modes for WLAN

1.4 MAC access channel

IEEE802.11 defines two basic techniques to access the channel, which are related to the mode of operation in the wireless network. Further, these techniques take place in MAC layer.

1.4.1 The Point Coordination Function (PCF)

PCF is a method to access channel in the infrastructure mode. It is a synchronous method to access the channel by using beacon frames at regular intervals, as described in figure 1.2. Thereby, the access point polls each station during the contention free period (CFP). Thus, there is no collision this period. Therefore, PCF can provide a time bounded service (QoS). Also, it can be used in combination with DCF to support a super-frame, as specified in figure 1.2 and figure 1.3. This super-frame has two periods, one for polling in the CFP and the other for DCF in the Contention Period (CP). Also in the infrastructure mode, all the STAs associated to the same AP have the same Basic Service Set (BSS). The BSS is composed at least of one AP and one STA that have a unique ID for association (BSSID). For the mobile STA, it can be associated to multiple AP. So, it is associated to Extended Service Set (ESS). This lets the mobile STA to change its AP by associating automatically with the AP that has the strong signal. And the name of this feature is Roaming.

However, we are concerned by enhancing the operation in ad-hoc mode not in infrastructure one. So, we described PCF in brief. On the other hand, we will describe DCF in details, which explains the operation of ad-hoc mode.

1.4.2 The Distributed Coordination Function (DCF)

DCF is a method to access the channel in a distributed way. It is an asynchronous method to access the channel randomly. It represents the process to access the channel in ad-hoc mode. Also, it is used in infrastructure mode as well as ad-hoc mode, as described in figure 1.2.

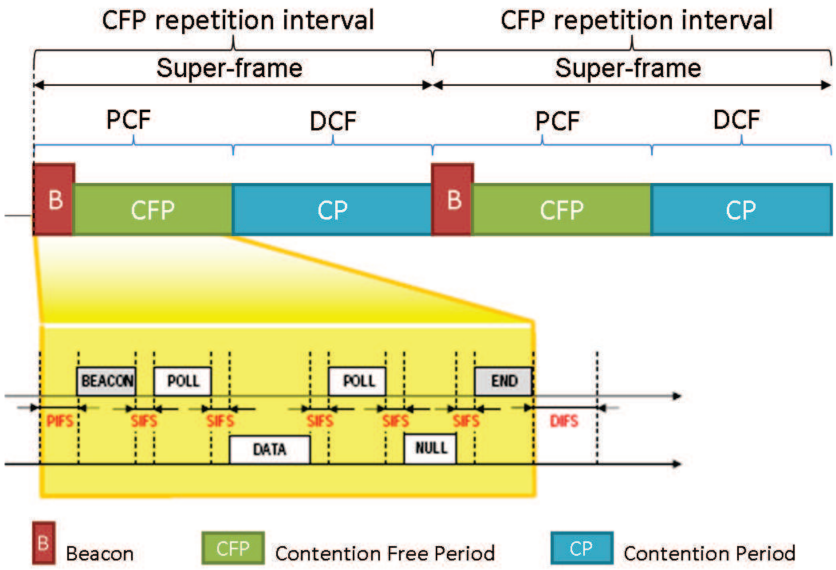


Figure 1.2: PCF access method

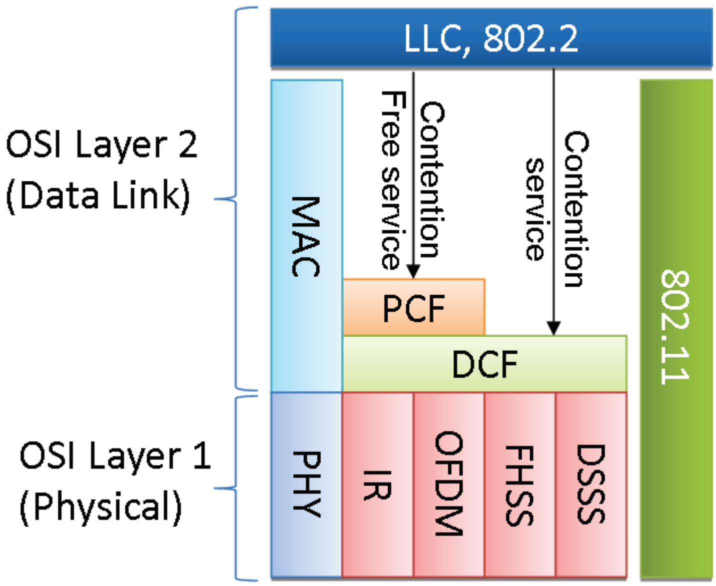


Figure 1.3: Layer function of PCF and DCF

We will describe this part in details because our studies are related to this part and its enhancements, all in ad-hoc mode. However, DCF does not support delay sensitive data transmissions applications (e.g. Multimedia, Real-time applications) since it is a simple technique designed to access channel for the Best-effort data transmission applications (e.g. FTP, Mail, Browsing internet). So, it cannot guarantee QoS.

However, the algorithm of DCF Access channel (Gast 05) is expressed in algorithm 1. This algorithm describes the functionalities of DCF in MAC layer to access the channel; channel free, channel busy or collision occurrence.

As explained in algorithm 1 and figure 1.4 (Gast 05), DCF is based on the Carrier Sense Multiple Access Collision Avoidance (CSMA/CA) with slotted time scale. The slotted time scale defined as the back-off counter. Thereby, every competing node attempts to transmit when its back-off counter is expired. Also, other time scales are defined as Inter-frame spaces. There are three inter-frame spaces in the original standard IEEE802.11. They are specified as following:

- SIFS: is the short inter-frame space, which is used for accessing the channel first to enable ACK, RTS, CTS or the highest priority transmissions (i.e. SIFS is equal to $10\mu s$ for the standard IEEE80.2.11b).

- DIFS: is the DCF inter-frame space, which is used for the contention based services (i.e. initiating the back-off counter) especially in ad-hoc mode or at the CP in infrastructure mode. However, when the channel is free within this period, each node may access the channel in contention based services (i.e. DIFS is equal to $50\mu s$ for the standard IEEE80.2.11b). In most of versions of IEEE802.11, DIFS is defined by equation (1.1). It can be verified as seen in table 1.3:

  $$DIFS = SIFS + 2 * SlotTime \tag{1.1}$$

  Moreover, DIFS is redefined to multiple values to classify and prioritize the data packets for QoS, as described in the EDCA, (see section 2.4.2).

Algorithm 1 DCF access channel

Require: Receiving_packet_from_upper_layer
    $attempt \Leftarrow 1; backoff \Leftarrow 0; CW \Leftarrow CW_{min}$
    Listen: Free_channel
    if channel is not idle **then**
        $attempt \Leftarrow attempt + 1$
        goto Listen
    end if
    Wait DIFS
    if $attempt \neq 1$ **then**
        if $backoff = 0$ **then**
            $backoff \Leftarrow Random\_value(0, CW)$
        end if
        Counting_down_backoff
        while $backoff \neq 0$ **do**
            $backoff \Leftarrow backoff - 1$
            if channel is not idle **then**
                goto Listen
            end if
        end while
    end if
    if channel is not idle **then**
        goto Listen
    end if
    $attempt \Leftarrow attempt + 1$
    Transmitting_packet
    Start ACK_timer
    if ACK_received = false **then**
        Collision
        $CW \Leftarrow CW \times 2$
        if $CW \geqslant CW_{max}$ **then**
            $CW \Leftarrow CW_{max}$
        end if
        if Exceeding_retransmission_number = true **then**
            Print ”Error of transmission”
        else
            goto Listen
        end if
    else
        Print ”Successful transmission”
        $CW \Leftarrow CW_{min}$
    end if

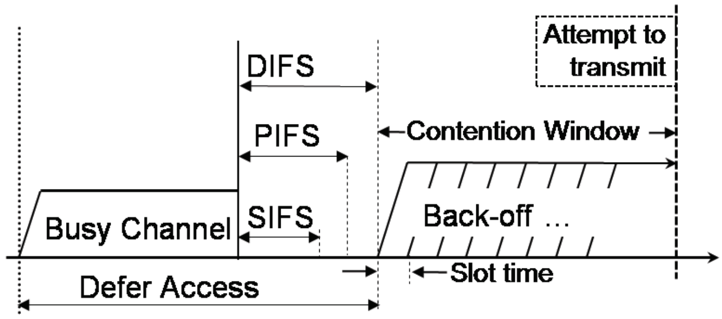


Figure 1.4: The Inter-frames for accessing the channel

- PIFS: is the PCF inter-frame space, which is used for the Contention Free Period (CFP) (i.e. initiating the beacon frame and the polling packet) especially in the infrastructure mode. However, when the channel is free within this period, the AP may initiate a CFP to access the channel (i.e. PIFS is equal to $30\mu s$ for the standard IEEE80.2.11b). As such, CFP has a priority over than CP. Thus, the waiting time of DIFS is bigger than PIFS.

So, the time of SIFS is less than PIFS whose time is less than DIFS. Thereby, shorter waiting time implies higher priority. Logically, a station, that is waiting DIFS to attempt to access the channel in CP, may stay longer time than an AP, that need to wait PIFS to send the beacon in CFP. Therefore, this AP has more priority than that station to access the channel. And, the same thing for the control packets that are waiting SIFS to be sent, they have the highest priority.

In addition, the slot time can be defined as a unit step of time measured in (μs) and determined by the version of the standard of IEEE802.11 (i.e. the slot time is equal to $20\mu s$ for the standard IEEE80.2.11b), see table 1.3. Thereby, when the back-off number is equal to (x), the time of the back-off is equal to the slot time multiplied by x, and so on.

As in CSMA (see figure 1.5), a source node listens to the channel for a period of time DIFS before starting a new transmission. If the channel is idle or free within that period, it will decrement its back-off counter until

Table 1.3: the times of the IFS for multiple versions of IEEE802.11

Standard	802.11b	802.11a	802.11g
Slot time	$20\mu s$	$9\mu s$	$9\mu s$
SIFS	$10\mu s$	$16\mu s$	$10\mu s$
PIFS	$30\mu s$	$25\mu s$	$19\mu s$
DIFS	$50\mu s$	$34\mu s$	$28\mu s$
CW_{min}	31	15	16
CW_{max}	1023	1023	1024

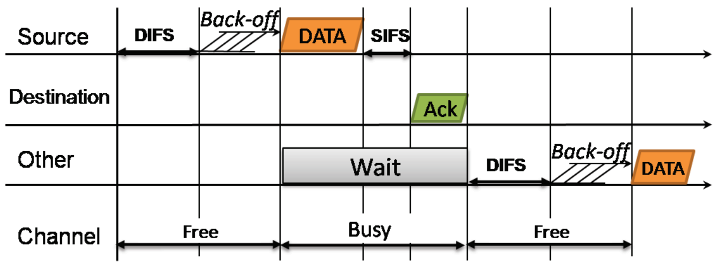


Figure 1.5: An example of DCF access channel

expiring. Thereafter, it will attempt for transmission. All the nodes around it are sensing that the channel is busy. So, they are entering in the wait state and suspending their back-off counter (defer access). The wait state is finished when the channel is returned idle by receiving an ACK message from the destination. Therefore, if there is another competing node that have to transmit, it will sense if the channel is idle for DIFS period. Then it will start counting down the back-off until expiring to attempt for transmitting the data packet.

However, this back-off is a random number of slot times. This number is chosen randomly and bounded between [0, CW]. CW is the contention window which starts from CWmin. The CWmin is one of the parameters to access the channel in DCF.

The example in figure 1.6 describes two competing nodes (Source1 and Source2). These nodes need to transmit at the same time after sensing that the channel is idle for DIFS period. In this example, for sake of simplicity,

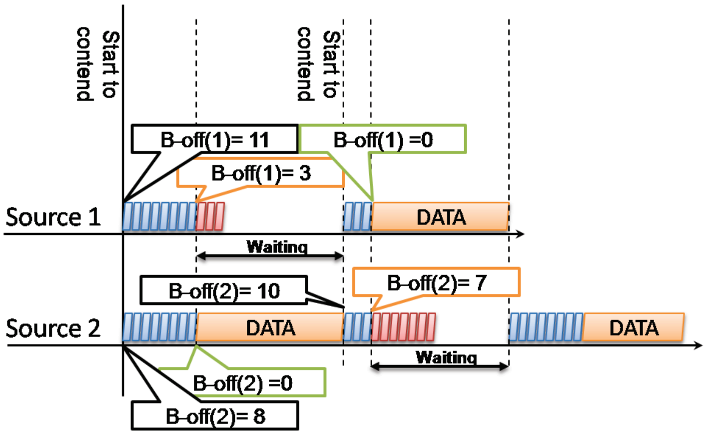


Figure 1.6: Competing to access the channel with neglecting to specify the control packets (i.e ACK, RTS/CTS, SIFS and DIFS)

we do not represent inter-frame intervals and control packets in figure 1.6. The first source node chose randomly its back-off (B-off(1)=11). The second one also choses its back-off (B-off(2)=8). Both of them start counting down their back-off. Thereby, the back-off of Source2 expires earlier. So, Source2 starts sending its data packet. At the other hand, the back-off of Source1 is suspended (B-off(1)=3). Then, after receiving the ACK message from the destination for successful transmission, Source2 chooses randomly a new back-off (B-off(2)=10) for new transmission. Both of them start counting down their back-off after DIFS period of idle channel. The residual back-off of Source1 will expire earlier after resuming decrement, and the back-off of Source2 is suspended (B-off(2)=7). Therefore, Source1 will transmit its data packet, and so on.

For the part of Collision Avoidance of (CSMA/CA) in the DCF, there is more than one treatment. The simplest one, as adopted in the original DCF, is doubling the CW (i.e. applying a binary exponential back-off algorithm) in the case of collision, as seen in equation (1.2).

$$CW_{new} = min(2 * CW_{old}, CW_{max}) \tag{1.2}$$

The CWmax is another parameter in DCF to access the channel, which is the maximum limit of the CW. However, after each successful transmission, the CW is reinitialized to CWmin.

At the opposite of the CSMA/CD (Ethernet), in wireless, the collision is avoided at the sender by the back-off mechanism and avoiding sending when the medium is busy. Thus, the collision happens at the receiver because of the hidden and exposed terminal problem, error check of CRC or channel error. These make the receiver abort sending the ACK. Therefore, the sender considers a collision is occurred after the timeout for ACK message.

The collision can be defined as; the moment for two source nodes attempt to send their packets at the same time. Because of the signal interference, the receiver cannot capture the correct signal. Therefore, the collide sources must retransmit their packets. Since the probability, that two source nodes will choose the same back-off number, is small, the probability of packet collisions, under normal circumstances, is low. So, this probability is perceptible when the number of source nodes is high.

Also for collision avoidance, an enhancement for DCF is based on (RTS/CTS) exchanging control frames to treat the hidden node or hidden terminal problem.

Therefore, each node has a transmission range (R), that is a circle of radius R, and this node is positioned in the center of the circle. The nodes covered in this circle can capture the signal of that node. However, not all the nodes are within the transmission range of each other. So, a hidden terminal problem simply can be explained by a small example of three nodes (A, B and C), see figure 1.7. A is within the transmission range of B, but out of the transmission range of C. While C is within the transmission range of B, C is out of the transmission range of A. However, A is transmitting to B. When C has a packet to transmit to B, C will sense an idle channel. Therefore C starts transmitting to B which causes a collision at B.

Therefore, to avoid the hidden node problem, (RTS/CTS) exchanging control frames are transmitted from the sender and the destination respectively, see figure 1.8. As in the last small example, when A attempts to

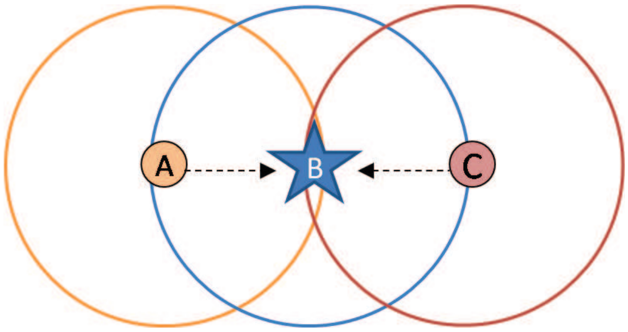


Figure 1.7: An example of the hidden terminal problem

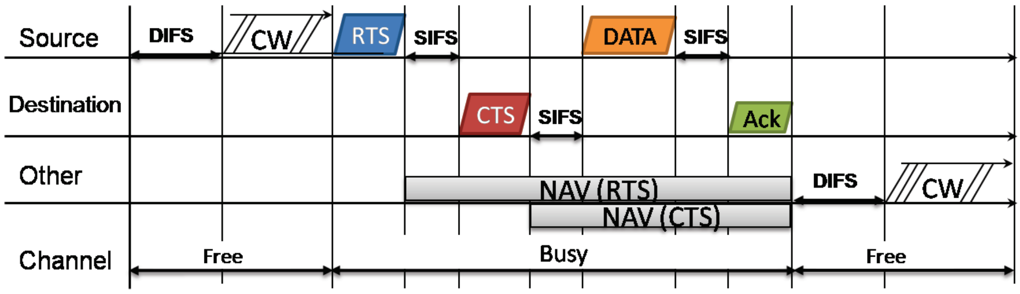


Figure 1.8: Avoiding the hidden terminal problem by (RTS/CTS) exchanging control frames

transmit its packet to B, it starts with sending the Request To Send packet (RTS) to B, then B transmits to A Clear To Send packet (CTS). Also, C receives this packet and waits NAV (CTS) until B transmits the ACK to A. Then C waits DIFS and attempts to transmit. So, all stations within the ranges of the sender and the receiver are suspending their transmissions by the network allocation vector, NAV (RTS) and NAV (CTS), when they hear the RTS and CTS respectively.

1.5 Conclusion

In this chapter, we had explained in details the principles, the technologies and important problems in wireless networks, specially ad-hoc WLAN and its functionalities. Thus, the popularity of WLAN encourages us improve some of its problematic sides, specially enhancing the QoS to guarantee the traffic in ad-hoc, and try to be overcome by our models (i.e. F-EDCA (Abuz 08; Abuz 09c) and FQ-EDCA (Abuz 09a; Abuz 09b)) with certain conditions. That makes the QoS enhanced and optimized.

Before presenting our models, we will present the state of the art and the main concepts of QoS in wired and wireless networks. These concepts will be described for local and end-to-end QoS applied in different layer levels (i.e. MAC layer and higher layers).

Chapter 2

Quality of Services (QoS)

QoS is a set of tools or technologies for managing network traffic in a cost effective manner to enhance the services for home and enterprise environments. QoS technologies have the abilities to detect changing network conditions (i.e. congestion traffic, bottlenecks, channel error, variable bandwidth, etc.).

These abilities allow prioritizing or throttling traffic. Thus, QoS technologies differentiate traffic and prioritize the critical and time-sensitive traffic over the non-critical traffic.

Definition 1. *(IETFa): the Quality of Service provides a manner that is designed to provide the service which is used to send the packets and is described with parameters, i.e. bandwidth, delays of packets and the losses of the packets.*

And also, it is highly recommended to provide a level of services similar to those available from the conventional wired networks. Thus, the services are classified in various classes depending on its service level.

Since the public network can carry many different types of flows, including voice, video, streaming music, web pages and e-mail, many of the proposed QoS mechanisms are complex or fail to meet the demands of the public Internet. Till now, there are data losses and packet delays. So, it is difficult to support each service with its requirements. For example, there is a critical traffic (i.e. audio and video streams and real-time data) that needs the guarantee for low latency and bounded delay. And there is a non-critical traffic (i.e. web pages, FTP, etc.) that does not require better than best-effort service.

Thereupon, QoS architecture introduces tools to treat packets differently by differentiating and classifying flows in categories or classes. Thus one flow receives better performance on the expense of others.

However, in the wired network side, there is a large research on optimizing QoS. For example, (Kadd 07) give results on measurements made on the Internet. These results show a good QoS in wired networks. Therefore, this enhancement of QoS could be considered not bad and better than the enhancement of QoS in wireless networks side. Although there is a lot of research concerning the QoS optimization in wireless networks, this optimization does not have at least the same level compared with that optimization in wired networks. This is referred to the problematic of wireless networks is not the same as that in wired networks. Thus, the wireless networks problematic leads also to face many difficulties for adopting and transforming the enhancement of wired side to wireless side.

In this chapter, we will present the QoS solutions, architectures, characteristics and managements for two parts. The first part is concerned for QoS in wired networks. The other one describes the QoS in wireless networks.

Indeed, the first part is necessary to show the ability to transform the protocols designed for wired networks and adapt them in wireless networks and their limits (ex. (Nian 02) describes an introduction of DiffServ QoS in 802.11 Standard). Thus, we will describe the most important architectures in wired networks (i.e. IntServ and DiffServ).

On the other hand, we will describe the QoS for wireless networks technologies, especially, that are designed for WLAN (i.e. the standard IEEE 802.11e). Because, there is a growing need to support QoS for wireless applications and mobile portable devices such as laptops, palmtops, etc. These devices become more and more popular, and the interest of using those portable devices without wires is growing very fast nowadays.

Further, this chapter could be described to introduce and summarize the QoS in three axes:

- The local QoS, that is referred to apply the QoS in hop-by-hop mode, against the global QoS, that is referred to apply the QoS in end-to-end mode.
- The QoS protocols applied in MAC layer level against those in higher layers levels, notably the network layer.
- The QoS approaches for wired networks against wireless networks.

Thus, by these axes, the QoS ideas and its conceptions almost are presented to describe the state of the art.

2.1 QoS Characteristics

The characteristics of QoS are considered as the parameters of QoS. Thereupon, the QoS is managed and controlled to serve the applications with guarantees and assured services. These parameters will be described as following:

2.1.1 The delay or latency

It is the time taken to transmit a packet from the application layer of the sender to the same level of the receiver. This time of period is called the end-to-end delay. It can be expressed as the summation of delays; sender process delay, propagation delay, buffer delay and receiver process delay.

However, the sender process delay could be made by the processing delay from the upper layer to the lower layer, and by the exchanged messages to establish a connection.

And also, the same thing could be for the receiver process delay, but in the opposite consequence. However, these different types of delays are categorized into two categories, as:

Constant delays

They include encoding and decoding times (for voice and video), and the propagation time, depending on the electrical or optical pulses speeds (Light speed, which is almost equal to 300.000 km/s in free space).

Variable delays

They refer generally to the network conditions, such as bottlenecks, queuing (scheduling) and congestion. That could be bounded or averaged to managing QoS.

So, the end-to-end delay could be bounded between the minimum delay and the maximum delay. Thus, the QoS can manage the resource allocations depending on the required service of the application.

2.1.2 Delay jitter

It is defined as the delay variance or as the difference time of the packets that arrive at the receiver in different sequences (packets have varied delays).

In other words, it is the difference of end-to-end delays between two sequenced packets. For example, two sequenced packets have end-to-end delay times; 50ms and the other 70ms. Thus, the delay jitter is 20ms. Therefrom, for veritable values, it causes distortion of voice signal.

2.1.3 Packet losses

Generally in networks, the packet is expressed as lost packet when it is dropped within queuing because of the congestion occurrences or bottlenecks. And also, when the receiver checks the CRC check error of a packet received, it find the packet corrupted because of signal degradation, saturated network links and noise interferences. So, the receiver neglects sending Ack to the sender. Therefore, the sender considers that packet is lost after the time-out of Ack waiting.

Further, it is defined as the difference between the total number of emitted packets and the total number of received packets.

2.1.4 Bandwidth

It is the allocated resource of the network. However, it can be measured by the rate of sending bits (the principle unit of data or packets). Also, this rate is variable (i.e. in Wireless networks) and depends on many factors, such as the technology (i.e. Ethernet 10/100/1000Mbps, WiFi 1/11/54Mbps), the topology, the media (i.e. fiber optics, wireless, twisted pair, coaxial, etc.), the interference, the routing protocols, processing, queuing, etc. And also, as bandwidth increases, power consumption increases. In a wireless device (i.e. hand-held device), this reduces battery life.

However, to guarantee QoS, global management of networks must be applied to control and manage the traffics of the networks, as described in the next section.

2.2 QoS Management

QoS management helps to set and evaluate QoS policies and goals. Further, the basic implementation of QoS in a network could be illustrated in three main componenets, as seen in figure 2.1.

Thus, supporting QoS in various networks requires collaboration of many components or techniques:

2.2.1 Admission Control

It limits the number of flows admitted into the network. So, each individual flow obtains its desired QoS specified by reservation resources, identification and marking techniques for coordinating QoS from end to end between network elements. Further, the policies to control and administer end-to-end

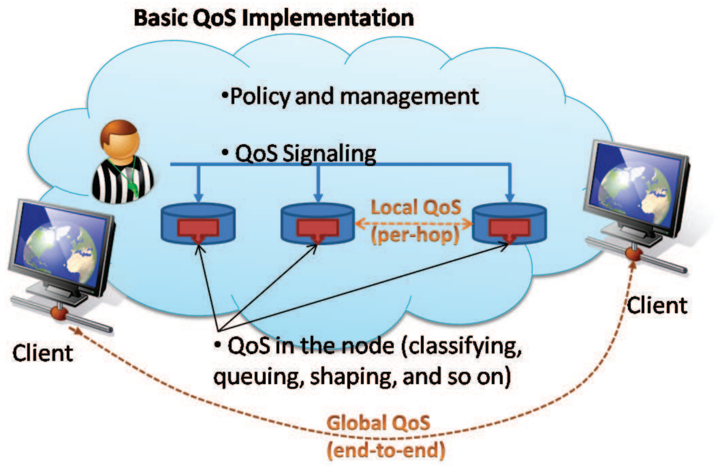


Figure 2.1: The basic implementation of QoS

traffic across a network, are specified by the Service Level Agreement (SLA), which is described in a lot of proposals like (Boui 02). Further, the most important protocol for admission control in the internet world is Transport Control Protocol (TCP) (RFC7 81b).

2.2.2 Classification

The classification of traffics provide preferential service to a type of traffic. So, to classify the traffic, it must be identified, while for marking traffic, it depends on treating traffic per-hop (local QoS) or end-to-end QoS. Thus, the traffic identified and not marked is treated for local QoS. While, the traffic identified and marked is treated for end-to-end QoS. As described, the identification and marking is specified by admission control. Therefore, the classification categorize the traffic to classes. Each class will be served differently depending on the class priority.

2.2.3 Scheduling and queue management

Generally, each traffic classified is composed of flows. The flow is expressed by condensed packets queued in each class. So, which packet is eligible and gets transmitted first on the output link? Significantly, it impacts QoS guarantees for different flows. Therefrom, it affects the parameters of QoS (ex. delay, jitter and loss rate).

However, there are a lot of scheduling algorithms or disciplines, such as; FCFS, GPS (Pare 93), WFQ (Kesh 89), etc. We will describe some of them later in the scheduling and queuing algorithms.

Therefore, some of these algorithms can be used to schedule the queued packets with fairness, weights or priority; depending on the packets' flows or packets' classes and the architecture model of QoS.

Queue management controls the queue size and decide which packets to drop during congestion. Therefore, it controls packet loss rate. However, managing queues is important to schedule data flows and to minimize and treat the congestion.

Thus, there are a lot of packet drop strategies such as (Fero 00), (Wurt 02), etc. The last one is Weighted Random Early Detection (WRED), which is one of the variety of RED and one of the congestion control techniques.

2.2.4 Congestion Control

It avoids, prevents, handles and recovers the network congestion. Therefore, a lot of proposals treat the congestion, such as; (Kesh 91), RED (Floy 93), WRED (Wurt 02), RIO (Clar 98). The last two proposals are extensions of RED (Random Early Detection) for challenging feedback problem (implicit feedback via dropped packets) with supporting QoS.

RIO is RED with In/Out, which supports by differentiating the input and the output flows. However, Internet traffic shows self-similar behavior in the TCP/IP networks that leads to the global synchronization.

Also, there are many of implementations for congestion control over TCP such as (RFC2b; Brak 95).

Further, the explicit congestion notification (ECN) (Rama 01; Floy 94) passes a congestion information (one bit of ECN is implemented in IP header) back to senders when the packet is dropped by the queue management.

However, we will describe the technologies and the solutions of QoS, as presented in the next sections. These technologies apply the parameters of QoS and attempt to guarantee the required services.

Therefore, let us present the state of the art for the QoS architectures in both of wired and wireless networks. We will describe the most important architectures for QoS, which are used nowadays, such as IntServ that is a per-flow model, and DiffServ that is a per-class model.

Furthermore, we will describe the compatibility of these architectures to adopt them in wireless networks and realize the required QoS (IEEE802.11e standard; HCCA and EDCA).

2.3 QoS in Wired networks

IntServ and DiffServ are QoS architectures, which are the most applicable nowadays in various networks. We will describe them in details.

2.3.1 IntServ (Integrated Services)

IntServ (RFC1) is a flow-based mechanism. This mechanism treats the traffic flow that belongs to the same source application. Where the end hosts signal their QoS needs to the network.

Then, IntServ follows the signaled QoS. The idea of IntServ is that; every router in the system implements IntServ path map, see figure 2.2. Thus, every application that requires some kind of guarantees has to make an individual reservation.

However, to implement the path map, the Flow Specifications describe what the reservation is for by two parts:

- Traffic Specification (TSPEC): this includes token bucket parameters.

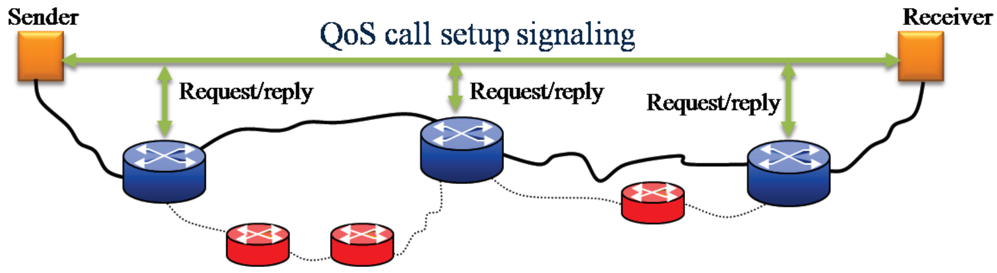


Figure 2.2: IntServ Architecture

- Request Specification (RSPEC): this specifies what requirements there are for the flow.

While Resource reservation protocol (RSVP) (Brad 97) is the underlying mechanism to signal the reservation resources across the network to create the path between sender and receiver.

The operability of the IntServ is to specify a fine-grained QoS system Resource reservation per flow using RSVP. Thus, IntServ must reserve the path map for each application flow in each router in the network. This makes installing IntServ in large networks is complicated and missing the scalability.

In the other hand, it can, for example, be used to allow video and sound to reach the receiver without interruption; similar to ATM virtual connections.

Further, this technology needs non interrupted connections or channels. Thus, it misses the efficiency when it is installed in the wireless networks. Because, in wireless networks especially Ad-hoc, the channel is not stable and the bandwidth is variable, which makes eventually the resource reservations not guaranteed. However, as described in (Chan 07), it adopts the context transfer protocol (CXTP) (RFC4 05) on IntServ wireless networks with infrastructure networks to resolve service interruptions during handoffs. Further, CXTP is used particularly for WPAN (i.e. Bluetooth) not for WLAN or 3G cellular networks.

Therefore, differentiating flows in classes may solve the problem of IntServ, as we will see in DiffServ.

2.3.2 DiffServ (Differentiated Services)

DiffServ (RFC2a; Cisc 05; Auri 04; Loch 04) is a class-based mechanism. It is a coarse-grained control system aggregation of flows into Per-Hop Behavior (PHB) groups. So it is a mechanism that treats data traffic separately in aggregate behaviors. That make the DiffServ to be scalable not like IntServ.

Its principle works on the provisioned QoS model, where network elements are set up to service multiple classes of traffic with varying QoS requirements.

Each router on the network is configured to differentiate traffic based on its class. Each traffic class can be managed differently, insuring preferential treatment for higher priority traffic on the network. The DiffServ model does not make judgment on what types of traffic should be given priority treatment since that is left up to the network operator (SLA).

Thereupon, DiffServ classifies flows packets, which allows serving traffic to be partitioned into multiple priority levels or classes depending on the flow requirement.

DiffServ Model (Domain)

A DiffServ domain consist of two parts; the edge routers and core routers, see figure 2.3:

The Edge routers (Boundary routers): They perform the classification and conditioning. A packet classifier classifies the flows packets depending on the packet information via Behavior Aggregate (BA) classifier. That classifier is searching the DSCP (Differentiated Services Code Point) (Nich 98) of the ToS (Type of Service, 8bit) in IP header, see figure 2.5. Thus, the classifier classifies the packets depending on the priority level of ToS with respect to the operator conditions TCA (Traffic Conditioning Agreement), which is part of the network provisioning contract of the service provider.

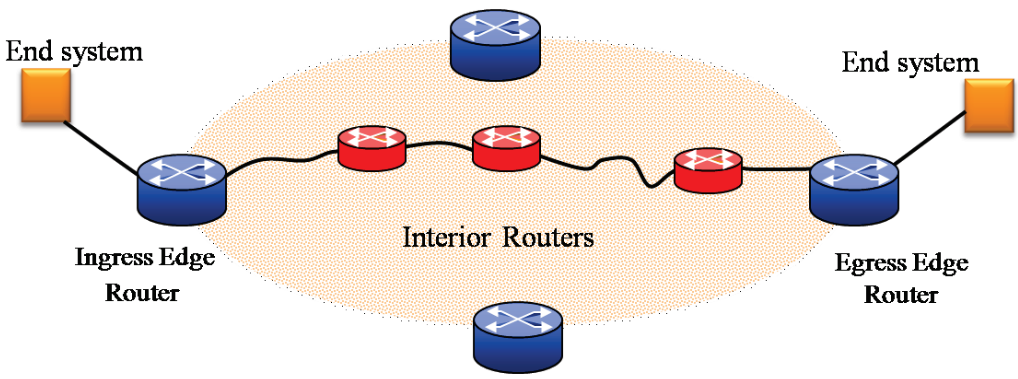


Figure 2.3: DiffServ Domain

However, when a packet is classified, it is forwarded to traffic conditioner that consists of Meter, Marker, Shaper and dropper, see figure 2.4.a. The meter measures the packet stream based on TCA profile. And also, the marker is marking the packets depending on its service by setting the DSCP value given by Classifier and Meter. Then, the Shaper is a traffic shaping technique for forwarding, queuing or dropping the packets depending on the policy of the temporal traffic conditions.

The Core routers (Interior): As the packet leaves the Ingress router and enters into the network core, PHBs are enforced to follow the traffic resource allocations per link based on the priority that defined in DSCP and packet marking.

The traffic resource is allocated and determined depending on the parameters of QoS, as explained; the packet loss rate, the delay and the jitter. Then scheduling and queuing disciplines (i.e. Weighted Fair Queuing (WFQ), Priority Queuing (PQ) and FIFO techniques) are forwarding the packets depending on the traffic resources available.

However, to avoid the congestion, many of techniques (i.e. WRED) treat this problem by dropping packets depending on their priority and traffic conditions, see figure 2.4.b.

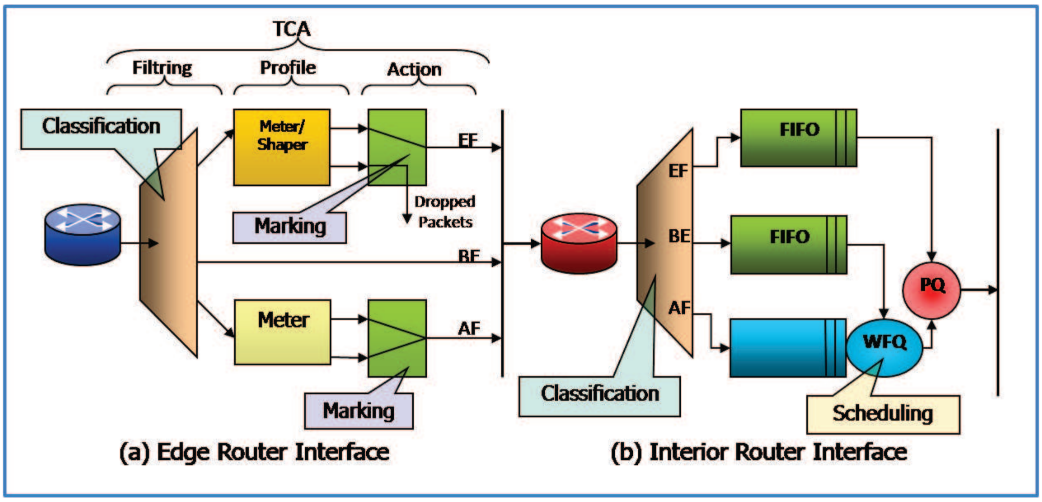


Figure 2.4: Architecture Model

DiffServ Architecture Field (DS Field) (Nich 98)

The packets can be marked using the DSCP in the edge routers with respect of the SLA. Then these packets are forwarded, buffered or dropped by PHB with respect to its level.

However, the packets that have the DSCP are aggregated with BA. Theses packets, which are sent from multiple applications or sources, could belong to the same BA. The definition of a class is expressed also by a BA. So, with DiffServ, QoS can be classified up to 64 different BA or classes.

These classifications are evolved from the 6 bits (2^6) of DSCP, which is placed in the ToS octet of IPv4 or the Traffic Class Octet of IPv6, see figure 2.5.

However, the QoS levels used nowadays with respect to SLA are as the following:

- Best effort service (BE): This is the service without QoS that belongs to time-insensitive applications. It is essentially for packet switched networks (ex. Internet).

- Expedited Forwarding (EF): It provides a hard guaranteed service to

Table 2.1: Assured Forwarding (AFxy) levels, with its DSCP

DROP Precedence	Class 1	Class 2	Class 3	Class 4
Low Drop Precedence	(AF11) 001010	(AF21) 010010	(AF31) 011010	(AF41) 100010
Medium Drop Precedence	(AF12) 001100	(AF22) 010100	(AF32) 011100	(AF42) 100100
High Drop Precedence	(AF13) 001110	(AF23) 010110	(AF33) 011110	(AF43) 100110

end users. Thus, it is applied for the critical and time-sensitive application. Further, its parameters of QoS are deterministic by a value (ex. Delay $<$ 10ms) or bounded values (ex. Delay [10ms, 80ms]).

- Assured Forwarding (AF): It is expressed as better than best effort. It provides a guaranteed service (with different drop precedence) to end users. And its parameters of QoS depend on predictivity, probability (ex. Delay $<$ 10ms with 75%) and stochastic distribution. This service is supported for the applications of priority level between the EF and BE.

Moreover, as seen in table 2.1, the AF could be able to classify the packets in 12 priority class described as (AFxy) referenced with DSCP.

Thus, it defines a method by which PHBs can be given different forwarding assurances. The AFxy in PHB defines four AFx classes: AF1, AF2, AF3 and AF4. Each class is assigned a certain amount of buffer space and bandwidth, dependent on the SLA with the Service Provider/policy.

For example, traffic can be divided into gold, silver, and bronze classes. The gold class allocates 50 percent of the available link bandwidth, silver 30 percent and bronze 20 percent.

Also, within each AFx class, it is possible to specify three drop precedence values when there is congestion in a DiffServ Domain on a specific link.

As seen in figure2.5, the DSCP is 6-bit field in the IP header and is defined in RFC 2474 (Nich 98). The first six bits contain the DiffServ Code Points

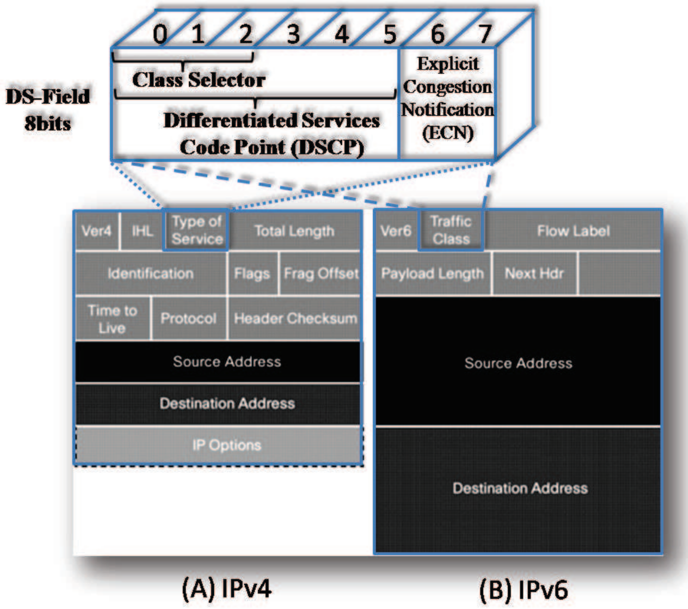


Figure 2.5: DSCP definition with ECN in IP header (A) v4 and (B) v6.

(DSCP). The two least significant bits (LSBs) of the DiffServ field (bits 6 and 7) are used for Explicit Congestion Notifications (ECN) as defined in RFC 3168 (Rama 01) and usable for TCP flows.

DiffServ has a skeleton of how to implement efficient service differentiation. And also, it needs to be carefully studied as how to efficiently manage network resources to provide powerful service in practical networks.

As a consequence, this could be useful to adopt the service differentiation technique in another world of networks, such as Wireless Networks (i.e.IEEE802.11e), and to support QoS.

We will present a comparison between DiffServ and IntServ, as in table 2.2.

In addition, in wired LAN, the DiffServ model allows controlling QoS requirements at the network layer level. DS-TE approaches such as PEMS (Lee 06) are techniques that give an end-to-end control of QoS requirements.

The challenge for WLAN and notably ad-hoc network is to be able to apply a DS-TE like technique at the link layer level.

Table 2.2: Comparison between IntServ VS. DiffServ

Comparison	IntServ	DiffServ
Packets classification	flows	classes
Granularity of service	individual flow	aggregate flows
State in routers	per flow	per aggregate
Traffic classification	several header fields	DS field
Signaling protocol	required(RSVP)	not required
Flow aggregation	fine-grained	coarse-grained
Coordination of service	global (End-to-End)	local (Per-Hop)
Priority service	for each flow	for each class
Scalability limited	by number of flows	by number of service classes
Network management	as circuit switching networks	as existing IP networks

However, we will present also the state of the art in wireless networks.

2.4 QoS in Wireless

This section describes the quality of service in the environment of WLAN implementations. The QoS guarantees in WLAN refer to its capability to support differentiated services for the network traffics.

These services are prioritized depending on the sensitivity of time, which called the QoS parameters, of those traffics. Therefore, these traffics can be managed to serve the higher priority with greater bandwidth and less delay than the lower priority with lower bandwidth and more delay.

And also, the bandwidth can be managed efficiently by controlling the network congestion and traffic shaping. This make the capability for controlling the jitter and the delay of traffics is more efficient.

The problems of wireless networks are characterized by many features, that affect the traffic link. They are Channel-error and Location-dependent as described in (Lu 97). The channel-error causes high error-rate and bursty-error, while, location-dependent is caused by user mobility. So, a high error-rate makes the bandwidth of the network is scarce and the mobility makes the bandwidth of the network is variable (time-varying). Thereby, a time-varying wireless link capacity and scarce bandwidth limit QoS guarantees and make the development of QoS hard challenge in wireless networks.

However, it is highly recommended to provide a level of services similar to those available from the conventional wired networks. Thus, the services are classified in various classes depending on service level.

In the last decade, the market of wireless networks is grown rapidly while the prices of its devices are decreased. This makes the interest and the popularity of the wireless networks, especially those are based on IEEE 802.11 (IEEE 97), is also increased. Thus, IEEE 802.11 wireless LAN has become a dominant broadband wireless technology in recent years. Therefore, the development of WLAN is crucial and important.

The wireless applications that need Quality of Services guarantees are growing too. These applications have time-sensitive data such as real-time data, voice and video. Their traffic data, which should be associated with these services classes, must be delivered to those wireless portable devices at specific data rates and within specific delays, packet loss and jitter bounds, which are the parameters requirements of QoS.

Moreover, at the beginning of developing the wireless network, it was mainly used for the best-effort and the low bandwidth data traffics. That means there are no QoS guarantees. Nowadays, the technologies of wireless networks are required to transport the critical and time sensitive data traffics, and also to support in various environments. This requirement led to the necessity for QoS in wireless networks. Thus, QoS is one of the challenges, which must be overcome, to realize the practical benefits in wireless networks.

Therefore, the upcoming IEEE 802.11e standard (IEEE 05a) enhances the original IEEE 802.11 MAC layer to support applications with QoS requirements by introducing Hybrid Coordination Function (HCF), which includes two techniques for medium access:

- Enhanced distributed channel access (EDCA): that administrates QoS in each station. EDCA is a distributed technique and can be used for ad-hoc mode. This is the architecture model concerned for. Further, we will focus on this model.

- HCF controlled channel access (HCCA): that administrates QoS in the base station (QAP). HCCA is a centralized technique and manages access to the medium using a QoS access point, which can be used for the infrastructure mode.

In IEEE 802.11, PCF has a poor QoS performance. Therefore, it does not define classes of traffic. And also, it uses a simple round-robin algorithm, which cannot handle the various QoS requirements. Moreover, the transmission time of the polled stations is unknown, that make the station to reserve the channel until finish sending. Thus, the problems of PCF are resolved by HCCA.

On the other hand, in ad-hoc, DCF has also a poor QoS performance, because it is designed for equal priorities, and the station may keep the channel as long as it needs.

Therefore, HCF provides guaranteed services. Whereas, it is enhanced by providing delay guarantees in IEEE 802.11e Networks (Anne 04). However, it combines the advantages of PCF and DCF. And also, it coordinates the traffic in many scheduling mechanism (not only round-robin). So, both EDCA and HCCA define Traffic Classes, to classify packets depending on their priorities.

2.4.1 HCCA

HCCA enhanced the PCF that is used in infrastructure mode, by developing scheduling mechanism in the base station (AP) that separates the flows in each queue depending on their priority. That means, it classifies the packets depending on its type (best-effort, video and voice, etc.).

However, in our studies, we are concerned for optimizing the QoS in ad-hoc mode. So, all the detailed aspects of the HCCA are beyond the scope of our studies as we focus on the HCF contention-based channel access (i.e. EDCA). Thus we will present EDCA.

As well, QoS over the wireless link is achieved via admission control, resource allocation and packet scheduling at the base stations (AP) in infrastructure mode. Even so, it is not the same QoS administration over the wireless link in ad-hoc. Indeed, in ad-hoc, there is no admission control that manages all traffic for all hosts. Every host station manages its data separately, which makes QoS in ad-hoc is a hard and complicated challenge. Therefore, in ad-hoc, it cannot guarantee QoS for all conditions and for all the time.

Thus, we need to treat each condition separately, which leads to enhance QoS guarantee in ad-hoc wireless networks.

Let us explain the architecture of QoS in wireless ad-hoc mode (EDCA).

2.4.2 EDCA

EDCA enhanced DCF that used in ad-hoc mode. The principles of DiffServ for differentiating flows is adapted and translated in EDCA by introducing the concepts of access category (AC) (IEEE 05a). It provides service differentiation and classifies the traffic of 8 different user priorities , as defined in three bits of TOS in the header packet of the upper layer (i.e. Network layer), see figure 2.5.

Thereby, each data packet from the upper layer with a specific user priority value should be mapped into a corresponding Access Categories (AC), see table 2.3. Therefrom, the user priority value is defined in the IEEE 802.1d bridge specification (IEEE 98a).

However, each station has four ACs to provide service differentiation, see figure 2.6. Moreover, it implements an Arbitrary Inter-Frame Space (AIFS). AIFS is a waiting time before the beginning countdown of back-off procedure. So, this waiting time is equal to a number of slot-times. This number is referred to AIFSN, which is a parameter in EDCA to access the channel.

Every AC behaves as a single DCF contending entity, and each entity has its own contention parameters (i.e. $AIFS[AC]$, $CW_{min}[AC]$, $CW_{max}[AC]$), see table 2.3. Thereby, each AC supports different kinds of applications such as; background data, best-effort data, video data and voice data, which each traffic data can be mapped to different AC queue (i.e. AC_BK, AC_BE, AC_VI, AC_VO respectively).

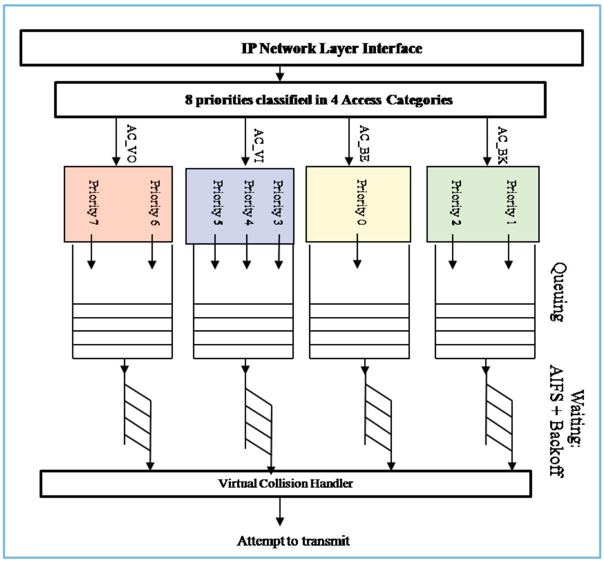


Figure 2.6: The architecture of EDCA

As seen below, each AC corresponds implicitly to a priority level to access the medium and provides guaranteed services.

- AC_BK is the lowest priority for background data.
- AC_BE is the next priority for best-effort data.
- C_VI is the priority for video applications.
- AC_VO is the priority for voice applications.

The first category is specified for the data that consumes the most of bandwidth and is not a time-sencitive (i.e. FTP). The second one is for the internet (IP) data. The last two access categories are the highest priorities reserved to classify the packets of multimedia applications.

Thereby, the packets, that have the most priority, access the channel first by controlling the parameters of EDCA (i.e. $AIFSN[AC]$, $CW_{max}[AC]$ and $CW_{min}[AC]$) to small values, see table 2.3. So, the smaller values of these parameters have the shorter channel access delay for the corresponding AC, and thus the higher priority to access the medium.

Table 2.3: The values of parameters of EDCA assigned for each user priority that mapped for each AC.

AC	**Traffic Type**	**User Priority**	**AIFSN**	CW_{min}	CW_{max}
AC_BK	Background	1	7	31	1023
	Background	2			
AC_BE	Best Effort	0	3	31	1023
AC_VI	Video	3	2	15	31
	Video	4			
	Video	5			
AC_VO	Voice	6	2	7	15
	Voice	7			

As seen in table 2.3, AIFSN is equal or bigger than 2. Thereby, AIFS can be determined by (2.1). Thus, the lowest AIFS can accelerate back-off countdown and vice versa. So, AIFS is equal to DIFS, which is the minimum arbitrary waiting time for the highest priority.

$$AIFS(AC) = SIFS + AIFSN(AC) * SlotTime \tag{2.1}$$

For example, as seen in figure 2.7, let the packets that are classified in AC_x have a privileged priority more than those in AC_y. As a consequence, the time waiting for data of $AIFS(AC_x)$ is less than $AIFS(AC_y)$ to start deferring access channel. Then, data of AC_x attempt to access channel and the other categories wait and pause their back-off's counting down.

According to EDCA, it behaves as DCF scheme for each AC with an extra procedure of waiting (AIFS). Thus, after the classification, the packets for the same AC are enqueued in FCFS (First Come First Serve or FIFO) simple queue with drop tail technique (drop the last packet when the queue is full). Then the packet is dequeued when it can access the channel to attempt for transmission after a waiting time. This waiting time is the summation of

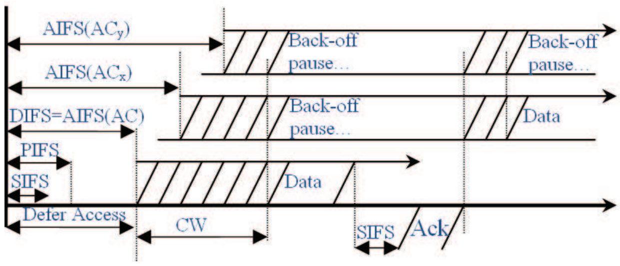


Figure 2.7: The Inter-frame access scheme for IEEE802.11e

deferring time of the AIFS[AC], depending on its AC, and the back-off[AC] countdown time. The back-off[AC] is chosen randomly between the interval [0, CW(AC)] for each AC.

Every AC decrements its back-off counter. When the back-off[AC] reaches zero, the node attempts to transmit the elected packet for that AC and the other categories wait and pause back-off's counting down. If the channel is clear or idle, then it transmits the packet and access the channel. Normally, the probability of packet collisions is high when the number of nodes is high.

However, the CW is reset to its minimum value ($CW_{min}[AC]$) after each successful transmission. Otherwise, it is doubled when a collision occurs or the medium is sensed to be busy at the end of defer access period, as seen in (2.2).

$$CW_{new}[AC] = min(2 * CW_{old}[AC], CW_{max}[AC]) \tag{2.2}$$

2.4.3 Other proposals for optimizing QoS in WLAN (RA-MAC)

IEEE802.11e and IEEE802.1D/Q (QoS standards for wired networks include IEEE802.1P) (IEEE 98a; IEEE 98b) are based on the second Layer (MAC). And also, the most of approaches, in wireless networks for QoS, enhances the QoS in the link layer level or MAC layer.

(Lu 97) proposes a new wireless link layer based on additional QoS parameters. They require frequent link state updates that are extra heavy loads, which can reduce the traffic quality in the network.

There are many of proposal that incorporate the QoS mechanisms of wired networks into wireless networks (Maha 01; Xu 05; Chan 07; Gall 07; Nian 02).

(Maha 01) proposes an experimental framework for supporting QoS in wireless networks using Differentiated services (DiffServ). However its results show that the need for a signaling and messaging protocol, which considers the variable bandwidth and mobility characteristics of the network, as opposed to the implicit admission control of DiffServ network. And also, the need for signaling and bandwidth allocation is important when the user moves between regions.

QoS based on applications suffer from interruptions during a wireless connection link because of an overhead caused by route discoveries that leads to an extra heavy traffic. Therefore, to avoid this, stability routing is needed (Gerh 03); resource constraints to support different demands for traffic classes that need DiffServ to achieve the optimal utilization. Unfortunately, DiffServ principles need the high layers (i.e. Network layer) to obtain the optimal utilization and guarantee QoS, although IEEE802.11e classifies the traffic in MAC layer.

Also, many of researches of QoS enhancement in ad-hoc wireless networks mainly focus on station-based DCF enhancement scheme such as (Nait 03; Bhar 94; Aad 01) or queue-based enhancement scheme such as (Mang 02; Romd 03; Wiet). Further, (Nian 04) enhance a schema of QoS in wireless networks

However, EDCA still has the problems of transmission and access channel such as hidden and exposed terminals problems (Kose 08). In (Zhai 05), many of proposal are proposed to solve with challenges the last problems, but this leads to new problems such as wasting bandwidth and channel utilization (spatial reuse).

Indeed, as described in (Chen 06), in EDCA there is no admission control that manages all traffic for all hosts. Every host station manages its data separately, which makes QoS in ad-hoc is a hard challenge (Zhai 05). In the other hand, it is not simple to adapt the QoS protocols of wired networks in wireless networks. Because the problems in wireless networks are not the same for wired networks such as a variable scarce bandwidth, channel error, etc.

As known, MAC protocol is based on a CSMA/CA scheme. Thereby, many of MAC protocols have been developed and improved to cope with problems that occur when a random access of channel, such as; MACA (Multiple Access with Collision Avoidance) (Karn 90) and MACAW (MACA with CW optimization) (Bhar 94). They are proposed to resolve the multi-access problems (Collisions) over wireless channels (mainly the hidden-terminal problem), and to enhance the channel performance and fairness.

Moreover, one of the proposals that update MACAW is Routing-aware Adaptive MAC (RAMAC) (Nait 03). Whereby, in wireless ad-hoc networks, data packets have to be routed hop-by-hop from a given source node to a destination node. This means that all of station nodes must participate cooperatively and accept to forward data for the benefit of routing of other nodes. If a number of nodes start to behave uncooperatively, the network may break down completely. So, a compensation of bandwidth of routing nodes has to be made in order to encourage them in routing other nodes' packets without any degradation of their own data transmission.

It has been shown in RAMAC (Nait 03) that this ability of forwarding packets leads to a new unfairness problem in wireless ad-hoc networks. Such that a node, which is forwarding other nodes' packets, gets less bandwidth, for its own packets, than a node which is not participating to the routing service.

The proposed routing-aware adaptive MAC (RAMAC) for legacy IEEE 802.11 shows all its effectiveness to cope with this unfairness problem in DCF.

However, to achieve fairness improvement, an extra bandwidth is allowed sometimes to the routing node's own traffic comparing to other non routing nodes' own traffics. Thus, differentiation between own and routed packets is needed in routing nodes.

And also, routing node has to access more frequently the channel than another node which is not participating in routing. Therefore, depending on the amount of data to route, CW is increased when there is a collision (after each unsuccessful transmission) by multiplying it by a multiplicative factor MF_i. By default, this factor is equals to 2 in DCF and EDCA as in (2.2). While, in RAMAC, MF_i is smaller than the value 2. So, the computation of the new value of CW that belong to node (i) is described in (2.3, 2.4, 2.5)

$$CW_{new}(i) = min(MF(i) * CW_{old}(i), CW_{max}(i)) \tag{2.3}$$

$$MF(i) = 2 - \rho(i) \tag{2.4}$$

$$\rho(i) = Wr_i(t)/(Wo_i(t) + Wr_i(t)) \tag{2.5}$$

Where:

- $\rho(i)$ is the ratio of routed packets in routing node i over the total packets.
- $Wo_i(t)$ is the amount of the own data traffic to send, which belong to node i at the time t.
- $Wr_i(t)$ is the amount of the routed data traffic to send, which belong to other nodes at the time t.
- (t) is the time of unsuccessful access channel (collision).

Thereby, $MF(i)$ is ranged between 1 and 2, since the ratio $\rho(i)$ is in the range between 0 and 1. Thus, the node which is sending its own traffic and participating in routing other nodes' traffic has the lowest contention window value so that it should have the highest priority to access the channel.

In addition, an enhancement of RAMAC is proposed by (Nait 04) for DCF, it proposes a smoothing multiplicative factor (β). β is equal to a constant value between [0, 1], involved in the computation of the new contention window value, see (2.6), in order to reach a better fairness in sharing the bandwidth. (Nait 04) describes that combining both the traffic differentiation and smoothing the CW increase function, has a significant impact on the medium access fairness between nodes.

$$MF(i) = 2 - \rho(i) * \beta \tag{2.6}$$

However, after each successful transmission, the CW is reset to its minimum value (CW_{min}), because it helps the node to access the medium with a high probability.

Moreover, the approaches described above are proposed for enhancing the QoS in MAC layer wireless ad-hoc networks. Whereas, other approaches for QoS provision are applied in network layer that are dependent or independent on the layer below (i.e. MAC layer), as described in the next section.

2.4.4 QoS routing in ad-hoc

There is a wide range of QoS requirements for ad-hoc (i.e. Mobile ad-hoc (MANET) (IETFb)), including throughput, delay, reliability or energy consumption. Most of QoS routing protocols involve trade-offs between these QoS metrics, as described in (Redd 06). As a result, several of these protocols adopt flexible mechanisms for tuning QoS variables to support various applications.

As described in (Jurd 07), the QoS for routing protocols aim to satisfy one or more performance goals at the network or node level. These goals include high reliability, robustness, low delay, high throughput and efficient energy. As described, the nodes in ad-hoc networks operate independently and have individual communication goals. So, generally, QoS protocols in ad-hoc networks should be provided on the node level. Therefore, in ad-hoc networks the QoS guarantees must be enabled for each node and even for each connection to optimize the network performance.

The goal of routing protocols is to determine a good path (sequence of routers) through the network from the source to the destination and to optimize the use of network resources. Typically, a good path means minimum cost path that characterized by the delay, the congestion level, throughput, stability of routes, signal strength, link cost, etc. Whereas, wired networks have different cost characteristics. Thereby, in ad-hoc, after receiving a service request from the sender, the first task is to find a suitable loop-free path from the source to the destination that have the necessary resources to meet the QoS requirement of the desired service. Therefore, this process is known as the QoS routing.

Moreover, the traditional routing algorithms depend on two principles: distance vector or link state. Distance vector establish a good path depending on periodic exchange of messages and selecting the shortest path if several paths available. While, link state notifies all routers periodically about the current state of all physical links. Therefore, complex computation is needed for router to get a global view of the network.

However, in ad-hoc, every node can send its packets to its neighbors. And also it must be treated as a router for routing and forwarding the packets of their neighbors. Therefore, the path is constructed by the collaboration and the connectivity between nodes.

Thereby, an example illustrates the QoS routing in ad-hoc wireless network shown in figure 2.8. The numbers between the parentheses are the attributes of each link. These numbers represent the available bandwidth in Mbps and the delay in ms. Suppose a service required to send a data flow from node A to node D with an end-to-end bandwidth guarantee of 3Mbps and an end-to-end delay of 14ms. Whence, the QoS metrics can be classified as additive metrics, concave metrics and multiplicative metrics. Therefore, the cost path of bandwidth is a concave metric, that means it is calculated by the minimum cost link of bandwidth. While, the cost path of delay is an additive metric, that means the end-to-end delay cost is calculated by the summation cost link of delay.

Thereby, QoS routing searches for a path that has sufficient bandwidth and bounded delay to meet the QoS requirement of the flow. The possible

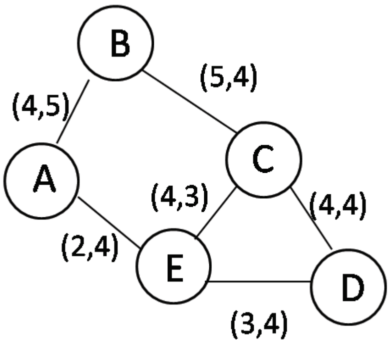


Figure 2.8: An example of QoS routing in ad-hoc

Table 2.4: The possible routing paths between nodes A and D

Path number	Path	N° Hops	BW cost (Mbps)	Delay cost (ms)
1	AED	2	2	8
2	AECD	3	2	11
3	ABCD	3	4	13
4	ABCED	4	3	16

paths available between nodes A and D are shown in table 2.4. As a result depending on the required service, the best path is the third. Although the first path has the minimum hops count and the forth path has a sufficient bandwidth, their costs are overpassed the service limit.

As described, QoS provisioning often requires negotiation between host node and network by call admission control and resource reservation by shaping and priority scheduling. However, the problematic of ad-hoc wireless networks such as the mobility, lack of resources, limited energy, no central coordination, etc., make the QoS of routing approaches challenging in ad-hoc wireless networks. Whereas, different applications have different QoS requirement. That means, each traffic application may follow different paths. Further, mobility make the topology of ad-hoc network is dynamically varying. That makes frequent paths breaks and discovers another paths. This re-establishment of paths may be unacceptable for some of application because of extra delay, packet losses, etc.

There are a lot of proposals of QoS routing solutions such as:

- DSDV (Destination Sequenced Distance Vector) (Broc 98; Perk 94)

- OLSR (Optimized Link State Routing) (RFC3 03b)
- PLBQR (Predictive Location-Based QoS Routing) (Shah 02)
- ZRP (Zone Routing Protocol) (Haas 97)
- CEDAR (Core Extraction Distributed Ad-hoc Routing) (Sinh 99)
- BR (Bandwidth Routing) (Lin 99)
- DSR (Dynamic Source Routing) (John 96)
- AODV (Ad-hoc On-demand Distance Vector) (Perk 99)
- QoSAODV (QoS enabled Ad-hoc On-demand Distance Vector) (Perk 00)
- TBP (Ticket-Based Probing) (Chen 99)
- TDR (Trigger-based Distributed Routing) (De 02)
- OQR (On-demand QoS Routing) (Lin 01)
- AQR (Asynchronous QoS Routing) (Vidh 03)
- OLMQR (On-demand Link-state Multi-path QoS Routing) (Chen 02)

As presented in figure 2.9, these protocols are classified on the network layer depending on the periodic update of routing state information. So, the state information, that are sent periodically, are called table-driven or proactive protocols. While, the state information, that are sent on-demand, are called on-demand or reactive protocols. Further, when the last two are implemented together, they are called hybrid protocols. Therefore, we will introduce some of QoS routing protocols that are implemented in network layer as following:

DSDV (Broc 98; Perk 94) is an expansion of distance vector routing. It is a table-driven protocol. Thus, each node keeps a routing table of destination, next-hop, number of hops and sequence number. The loop free is treated by the sequence number. The sequence number is originated by the destination

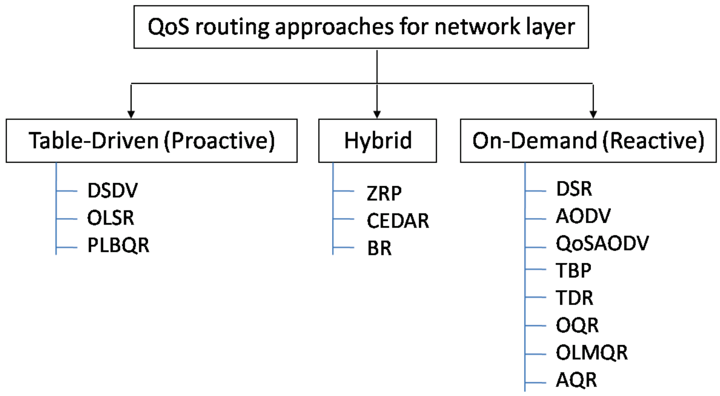


Figure 2.9: Classification of routing protocols for mobile ad-hoc networks

and indicating how is old the route. Every node manages its own sequence number by assigning it greater than the old one at every time of changing the route. However, each node maintains routing information for all known destinations. So, routing information must be updated periodically. That leads to traffic overhead even if there is no change in network topology. However, it is simple. And, there is no latency caused by route discovery and no dependency on the MAC layer. Thereby, it has best performance when node mobility rate and movement speed are low (i.e. stable topology). For that reason, DSDV is performed to discover the suitable loop free paths in the simulations of our enhancement (i.e. F-EDCA and FQ-EDCA) for EDCA on MAC layer.

On the other hand, another protocol built on DSDV is AODV (Perk 99). It is an on-demand protocol, because the update messages only get sent to nodes on a given route. Thus, route discovery floods all nodes with requests. A node responds the first time when it receives a request and replies only if it has a contact with the destination or has a valid route to the destination. However, it is similar to DSR (John 96) in route discovery, but it improves DSR by keeping routing tables of next hop.

Further, QoSAODV (Perk 00) is an on-demand (reactive) protocol. It extends the basic AODV (Perk 99) to provide QoS support in ad-hoc networks. Wherein, it includes Object extension on Route Request (RREQ) and Route

Reply (RREP) messages which specify QoS parameters (i.e. the bandwidth and the delay) during the phase of route discovery.

In addition, PLBQR (Shah 02) is a table-driven (proactive) protocol. This protocol is mainly to relieve the scalability issue with respect to communication overhead. Instead of distributing the state information of each link in the network wide. Each node broadcasts its current position, velocity, moving direction, and available resources on each of its outgoing links across the network periodically. Wherein, mobile hosts are equipped with Global Positioning System (GPS) receivers and their moving behavior is predictable. Therefore, at any instant each node can locally depict an instant view of the entire network. So, the source locally computes a path that accommodates a QoS request and route data packets along the calculated path. Further, the source can predict path break and predictively compute a new path before the old path breaks by using the stored global state information. This table-driven routing protocol is for providing soft QoS in small or medium sized networks, because it is inaccuracy in delay prediction.

Moreover, CEDAR (Sinh 99) is a hybrid protocol. It is designed to select routes with sufficient bandwidth resources. Therefore, it dynamically manages a core network, on which the state information of those links of stable bandwidth is incrementally propagated. Each core node is responsible for maintaining its local topology and calculating routes on behalf of nodes in its neighbors. CEDAR selects QoS routes upon request. Then, a core path is first established (from dominator) by flooding a control message across the core network. So, a QoS path calculation is then performed. However, this core path is shorten by using the partial information that are kept by the core. Thus, its performance depends on how well core nodes can manage their local resources. On the other hand, its disadvantages are specified as; the load factor or bottleneck at core nodes and how it can affect the network performance, are not addressed.

As a conclusion, QoS routing is a key issue in provision of QoS in Ad Hoc networks. However, no particular of QoS routing approaches provides an overall or a perfect solution, because these approaches have been proposed for focusing on different QoS metrics.

Although, QoS routing protocols could provide end-to-end QoS in ad-hoc networks, an integration and coordination with MAC layer is needed as described in (Zhan 05). Thus, as described above, DSDV is a traditional routing protocol and a principle of QoS routing protocols. So, this protocol is performed in the simulations (with fixed topologies) of our models (i.e. F-EDCA and FQ-EDCA) to enhance the QoS guarantees with fairness. Thereby, logically, these models could be integrated with a QoS routing protocol to enhance the QoS provision with fairness for mobile hosts in dynamic topologies.

Further, we will describe the technologies and the solutions of QoS depending on their layer level, as presented in the next section.

2.5 Layer QoS technologies

QoS functions in multi layer depending on the service or the structure of the networks, so we will present the QoS technologies depending on the layer OSI model, as the following:

2.5.1 Data Link layer technologies (Layer 2)

IEEE802.1Q/p/d (IEEE 98b; IEEE 98a) are the most important solutions for QoS in this layer. In the last decade, these standards dedicated for the Ethernet frames. Thereupon, the standard IEEE802.1p (included in IEEE802.1d) prioritizes the data packets encapsulated in the frame to 8 priorities. These priorities are specified by 3 bits in the header frame (i.e. Tag header). On the other hand, IEEE802.1Q is used to divide the performed LAN to multiple Virtual LANs (VLANs). Also, these standards are enabled by the announced Tag header in the frame header.

However, the solutions in that layer are considered as a QoS solution for LAN. And they could not be considered as a complete solution (End-to-end QoS) without taking into account the upper layers protocols and their specifications.

2.5.2 Network layer technologies (Layer 3)

Upon this layer, the end-to-end QoS solutions could be considered and evaluated. Recently, a lot of approaches are developed to enhance the QoS solutions. The main models are IntServ (RFC1) and DiffServ (RFC2a; Cisc 05) presented in section 2.3

2.5.3 Transport layer (Layer 4)

Port priority is the simplest way to indicate the priority for the type of data (ex. Port 21 for FTP low priority).

Although, TCP/UDP are the most famous protocols in the internet world, they could not guarantee QoS requirements for the upper layer (i.e. the application layer). Thus, Real-time Transport Protocol (RTP) (RFC3 03a) is developed and used extensively for real-time and media streams applications.

Further, there are some approaches applied for QoS enhancement in transport layer, such as SCTP (Stream Control Transport Protocol) (RFC2c; Stew 00) and AOTP (Application-Oriented Transport Protocol)(Tsao 00). AOTP is oriented for applications of upper layer and provide their requirements of QoS.

Thereupon, AOTP supports transport services, with new functionality added specifically, to trade off reliability, throughput and/or jitter. While, SCTP could be considered to meet the application requirement and translate them to the lower layers to achieve the end-to-end QoS. Further, another enhancement of SCTP applied in wireless environment (Ma 03).

These protocols are developed to replace the ancient protocols (UDP/TCP), because the last ancient protocols could not completely meet the strict requirements of the upper layer.

2.5.4 Application layer technologies (Upper layer)

The QoS technologies for the application layer, which map to lower layers, are concentrated on the process and new codec developing performance.

However, the most of time-sensitive applications, that need the QoS guarantees, are oriented for multimedia (i.e. audio and video), telephony (ex. VoIP), real-time, and so on.

For example, the image processing and video coding are developed for multimedia and video telephony (ex. standard H.264) and mapped to the lower layers protocols (i.e. Real-time Transport Protocol (RTP)/UDP).

Other approaches, like (Mulr 06), enhance the last development to degrade the congestion that leads to minimize losses and retransmissions. Thereupon, the QoS for these applications (simultaneous) is enhanced. And also, these applications could be synchronized, by scheduling. This leads to enhance the QoS.

Further, in (Bout 07), it describes a cross layer architecture for real-time applications. As a consequence, the environment of research for developing the QoS for multiple layers is large. So, a good distribution of tasks in each layer could simplify the QoS administrations and managements.

2.6 Traffic shaping and policing

Traffic shaping and policing (Evan 07) limit the full bandwidth potential of the traffic flows. They are used to avoid the overflow or the congestion. However the traffic shaping is using token bucket and leaky bucket schemes to optimize or guarantee performance, low latency and bandwidth, to deal with the concepts of QoS. So, these techniques regulate the traffic rate. While, policing is the same of traffic shaping, but it differs by discarding not buffering the traffic that exceeds its configured rate. However, the main concepts of traffic shaping are actually managed by the scheduling mechanisms.

However, traffic shaping schemes are commonly implemented at the network edges to control traffic that is entering the network (i.e. DiffServ).

Thus, upon identifying traffic flow using DiffServ, it applies a policy to control traffic's flows. Such that, some of the flows need either to guarantee a certain quality (i.e. Multimedia and VoIP) or to provide best-effort delivery.

This may be applied at the ingress router (the router at which traffic enters the network) with a granularity that allows the traffic shaping mechanism to separate traffic into classes or individual flows and shape them differently.

As a consequence, each flow could have the required resources depending on the service level agreement (SLA) and its policy. Further, it is necessary, for bounding the end-to-end delay, to satisfy the required components of QoS. Therefore, traffic shaping provides a limited and controlled rate when serving and queuing the packets in scheduling.

Actually, there are two methods for traffic shaping, which limit and control data rate of burst traffic. These methods could be adapted in QoS and scheduling disciplines:

2.6.1 Leaky Bucket

It is a method of traffic shaping that controls the incoming burst data traffic (Evan 07).

However, as seen in figure 2.10.a, if the incoming traffic with rate R, which is less than the bucket rate r, the outgoing traffic rate will be equal to R.

In this case, when it starts with an empty bucket, the rate of the incoming traffic is the same as the rate of the outgoing traffic as long as $R < r$.

And, as seen in figure 2.10.b, if the incoming traffic with rate R, which is greater than the bucket rate r, the outgoing traffic rate will be equal to r (bucket rate). That leads to extra packets will be stocked in the bucket, whose capacity is b (in bytes or in packets). As long as $R > r$, the bucket will be filled up, as seen in figure 2.10.c. Thereby, another extra packet will be dropped or treated differently depending on the policy.

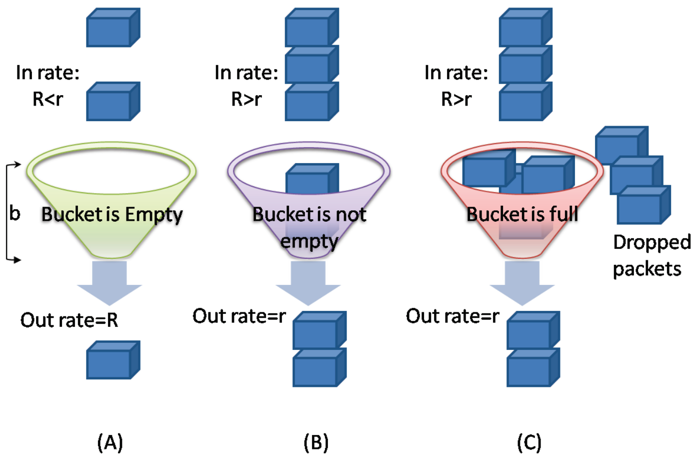


Figure 2.10: Leaky Bucket scheme; (A): the incoming rate is less than the leaky bucket rate, (B): the incoming rate is bigger than the leaky bucket rate and (C): the incoming rate is bigger than leak bucket rate with dropped packets.

2.6.2 Token Bucket

It is a method of traffic shaping that control the incoming rate traffic (Evan 07). It does not limit the traffic rate like leaky bucket, because its conception that there is a token for each forwarded packet. So, the rate flow depends on the token rate.

Thereby, as seen in figure 2.11.a, if the incoming traffic with rate R is less than the token arrival with rate r, the outgoing traffic rate will be equal to R. That makes extra tokens will be stocked in the bucket, which has a capacity of b tokens.

Thus, the rate of tokens is equal to zero when the bucket is filled up. Further, if R is greater than r, where the bucket is stocked of tokens, the outgoing traffic rate will be equal to R, as seen in figure 2.11.b.

As long as, R is bigger than r, eventually the stocked tokens will be exhausted, which will make the outgoing traffic rate is equal to r, see figure 2.11.c. In this case the incoming traffic has to wait for the new tokens to arrive in order to be able to send out. Therefore, the outgoing traffic is

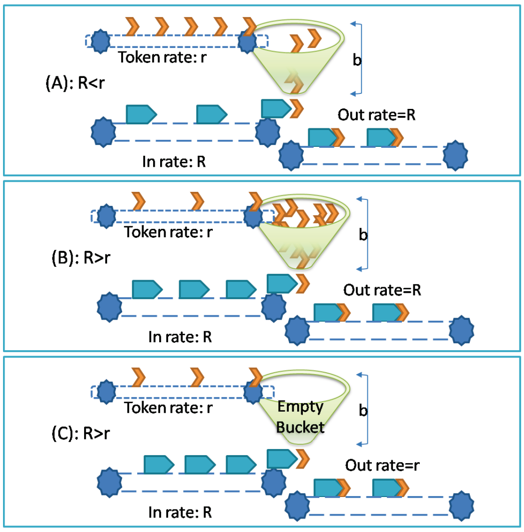


Figure 2.11: Token Bucket scheme; (A): the incoming rate is less than the token rate, (B): the incoming rate is bigger than the token rate with non-empty bucket, (C): the incoming rate is bigger than the token rate with empty bucket.

limited at token drained out rate.

However, it is very important for QoS architectures to adapting a method or mechanisms to arrange the flows in a fairly manner depending in its path or its class, which has the priority definitions.

So these mechanisms are suited for the scheduling and its main concepts. And also, some of these concepts are adopted in FQ-EDCA, as explained in chapter 4, by implementing a hierarchical scheduling mechanisms.

Thus, we will illustrate some of the recent mechanisms of scheduling.

2.7 Scheduling

Scheduling (Queuing) is a method to control flow in the buffer queues. Normally, implementing flow control is done in the source by controlling rate traffic of sending packets, or in the Gateway (Router) by Adaptive routing or Scheduling (queuing) the packets. It controls the order for forwarding the packets. It can be implemented to avoid congestion and to distribute the resource allocations of bandwidth equally (with Fairness) or weighted (for QoS). The gateway schedules packets from multiple flows and forward them for the same outgoing link (ex. Multiplexer, Round-robin). The simple solution for scheduling is that the packet which First come First served (FCFS) or FIFO that used for Best-effort service (without QoS). In the other hand, the scheduler that provide QoS guarantee, is a complex solution and statistical multiplexer for:

- Which packets should be given preference?
- How many packets should be transmitted from a flow?

Further, the scheduler that provides QoS guarantee is a hard challenge, because it depends on the type of network and its topology, the protocols of upper layers, the type of flows, etc.

The difference between scheduling with or without fairness (Simo 03), see figure 2.12. We will describe concretely the scheduler's mechanisms in this chapter. However, the packets have different size. Thus, when multiple flows are enqueued in FCFS queue, then the flows will be served with different allocations of bandwidth depending of their packets size. So, the packets that have the large size will consume the most of bandwidth. And also, when there are condensed packets enqueued in the queue, the other flows will wait these packets to be served. On the other hand, another story and treatment for the schedulers with different queues to treat each flow with fairness or guaranteed service.

However, the scheduling mechanisms can be classified in two conceptions (Klei 76; Lieb 99):

- Work-conserving scheduler: Always transmits a packet as long as there is at least a packet available in switch buffer. And also, it has an optimal performance in terms of throughput.

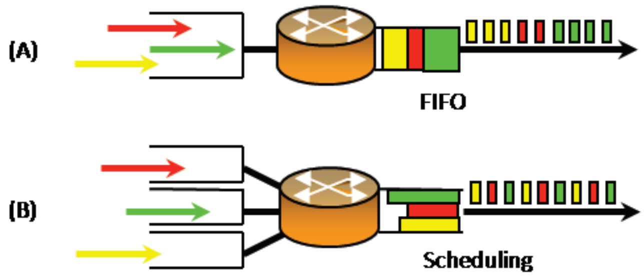


Figure 2.12: (A) Scheduling with the simple method FIFO, (B) Scheduling with fairness.

- Non-work-conserving scheduler: There is no transmission even if there are packets stored in buffers. This may delay packet transmission and reduce throughput. Though, it has better guarantees on delay jitter. However, it is in theory appealing approach, not much used in practice.

In addition, it is very important for bounding end-to-end delay to guarantee QoS requirements. Therefore, traffic shaping provides a limited and controlled rate when queuing and serving the packets with a proper scheduler discipline.

Actually, there are a lot of mechanisms of scheduling. Let us present the best known schedule algorithms:

- FIFO - FCFS (Mela 02).
- RR: Round Robin (Chas 99).
- WRR: Weighted Round Robin (Kate 91).
- SP: Strict Priority (Schm 04).
- CBSPQ: Class Based Strict Priority Queuing (Schm 04).
- PQ: Priority Queuing (Kenn 94).
- GPS: General Process sharing (Pare 93).

- WFQ: Weighted Fair Queuing (Kesh 89).
- SCFQ: self-clocked fair queuing (Gole 94).
- WF2Q: Worst-case Fair Weighted Fair Queueing (R 96).
- CBWFQ: Class-based weighted fair queuing (Cisca).
- LLQ: Low Latency Queuing (Ciscb).
- MMFS: max-min fair sharing (Zhan 96).
- VC: Virtual-Clock (Zhan 91).
- ADPQ: Adaptive Priority Queue (Khan 05).
- EDD: Earliest Due Date (Boda 00).
- DRR: Deficit Round-robin (Shre 96).

Generally, a lot of them have the same idea or the same characteristics for supporting fairness or guaranteed services for QoS, but till now, there is no perfect mechanism. We will describe in brief some of them:

2.7.1 First-Come-First-Served

FCFS or FIFO (Mela 02) is the simplest way of scheduling for Best-effort networking, see figure 2.12.a. The packets are queued into a common buffer. Then, they dequeued from front of queue. The first packet arrived, it is served first. However, this technique has a lot of disadvantages as; there is no flow isolation, which leads to unfair sharing of bandwidth. So a greedy (ill-behaved) source can occupy most of the queue and cause delay to other flows using the same queue, because there is no control for ill-behaved sources. Thus, TCP flows get penalized (congestion sensitive) with respect to UDP (no congestion control). And also, the time-sensitive flows (ex. Multimedia) will be served with degraded service. Therefore, there are no QoS guarantees.

And, it is hard to control the delay or bound it for the transmitted packets through FIFO buffers. In addition, for Congestion control, it uses drop-tail packet method; that means dropping the last packet arrive when the buffer is full.

However, FIFO is the principle architecture for EDCA, as described in section 2.4.2. Thereby, EDCA provides FIFO queuing method for each Access Category. Therefore, EDCA could miss fairness. For that reason, we will present a Fairness Queuing model (FQ-EDCA) that enhances EDCA, as described in chapter 4.

2.7.2 Round-robin

RR (Chas 99) is the simplest way of scheduling as a multiplexer for Best-effort networking, see figure 2.13. It has the same features of FIFO for the same queue. The difference is that; RR separates queues for each flow packets. And, the packets are served in a pure round-robin manner. However, there are disadvantages such as; it is unfair for different packet sizes. Theoretically, there is a conception, which is bit-by-bit round-robin that is impossible to apply. The scheduler serves a bit from each queue. So, each queue is served fairly. Practically, the scheduler serves a packet from each queue. Thus, RR is fair, if the packets have equal sizes. Therefore, Packet-by-packet Round-Robin (PRR) or Weighted Round-Robin (WRR) (Kate 91), that is the same of WFQ (Kesh 89), treats the packets of different sizes and solves the problem of unfairness.

In addition, there is no congestion-avoidance, which is caused by ill-behaved (mischievous) sources.

Actually, it is the principle method of scheduling that can be applied for QoS. And also, it is a work-conserving, which means that the server is not idle when there is a queued packet or non-empty queue.

However, the concepts of WRR are adopted in FQ-EDCA to apply fairness, as described in chapter 4.

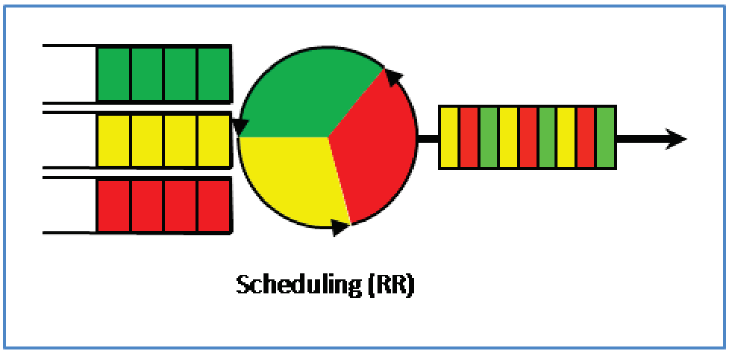


Figure 2.13: Round-robin scheduling

2.7.3 Priority Queuing (PQ)

PQ (Kenn 94) scheduler has multiple queues served differently with priority, and FCFS can be used in each queue. For example, multiple queues with priority (0) to (n). The queue with priority (0) is served first. When that queue is empty, the scheduler serves the queue below. That means, the queue of priority (i) is served, only if the queues of priorities from (0) to (i-1) are empty. Thus, the queue, which has the highest priority, has the lowest delay, lowest loss and highest bandwidth. Although QoS could be applied, a possible starvation of lower class may be occurring, because PQ is a Non work-conserving. That means there are packets for lower priority that are waiting the service.

However, the priority queuing in WLAN (EDCA) is adopted in a different way. EDCA supports priority queuing by waiting accessing the channel with different slot-times depending on the priority of the forwarded flow, as described in section 2.4.2. So, our models (FQ-EDCA and F-EDCA) are implemented in EDCA architecture. Thereby, the priority queuing in our models is already defined.

2.7.4 Strict Priority

SP (Schm 04) classifies flows into priority classes, which maintains only per-class queues and performs FIFO within each class. And the bandwidth is

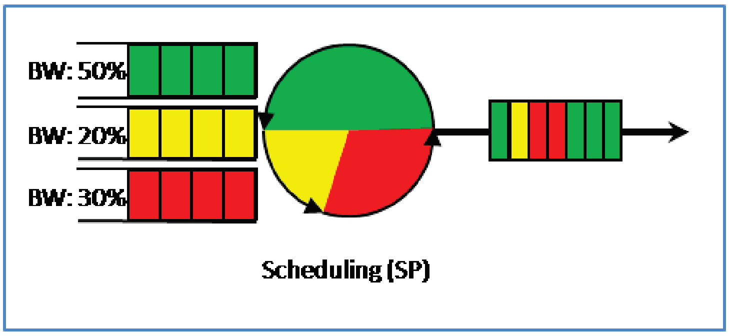


Figure 2.14: an example of SP scheduling

distributed for these classes. It can be explained by this example, see figure 2.14. There are 3 classes. Each class has a percentage of the bandwidth as; class 1 has a percentage 50% of the bandwidth and applied over the buffer queue that belongs to class 1, class 2 has a percentage 30% of the bandwidth and applied over the buffer queue that belongs to class 2, class 3 has a percentage 20% of the bandwidth and applied over the buffer queue that belongs to class 3.

Actually, SP is not like PQ. SP is treating with classes and PQ is treating with flows. Therefore, as explained in section 2.3, SP is applied in the DiffServ, because DiffServ treats classes. And, PQ is applied in the IntServ, because IntServ treats flows. And also, SP could avoid the starvation, because it is Work-conserving. As we will present in our models, we take into account these ideas and adopt them.

However, there is no significance for supporting a strict bandwidth (ex. 50% of bandwidth) for a source node in WLAN. Because the bandwidth in WLAN depends on accessing the channel. So, there is no guarantee to have that bandwidth. And also, the bandwidth in WLAN is variable because of the problematic sides in WLAN. Thus, a flexible bandwidth allocation technique with fairness is needed to enhance the QoS in WLAN, specially in ad-hoc.

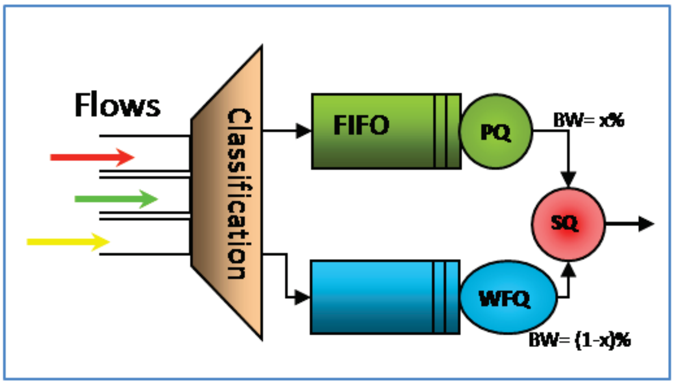


Figure 2.15: A model expresses LLQ and CB-WFQ architectures.

2.7.5 Low latency Priority and Class Based WFQ

LLQ (Ciscb) is a mechanism of scheduling to reduce jitter delay. It is a combination of the Priority Queuing (PQ) and the Weighted Fair Queuing (WFQ). It is designed exclusively for video and voice, which queued by strict queuing. And the other flows used WFQ, see figure 2.15. Also, LLQ is combined and expressed with CB-WFQ (Class Based WFQ) (Cisca).

As a conclusion, theses mechanisms are hard to be adopted in WLAN for IEEE802.11e standard. And also, they have a different architecture of that in EDCA. So, that makes it complex to transfer their concepts in EDCA.

2.7.6 General Process Sharing

GPS (Pare 93) is an efficient, flexible and analyzable discipline. Therefore, it seems very appropriate for integrated service networks. However, it is a fluid model scheduling mechanism. That means, the input traffic is infinitely divisible, and it serves the packet when it arrives. Thereby, it is not idle if the packet is waiting, because it is a work-conserving discipline. Moreover, GPS is a flow-based multiplexing discipline. So, it kepts flows in separate sessions. And also, all the sessions flows can be served simultaneously. Session is another definition of queued flow. Therefore, GPS use this term, because it is a fluid model that serves flow of infinitesimals not packets. While, the queue serves flow of packets.

In addition, GPS can ensure fair allocation of bandwidth among all backlogged (the queue is not empty) sessions. Thus, it is suitable for feedback based congestion control algorithms. In other words, it provides fair share services. Thus, all flows sessions have the same resources (the same slope for all flow session graphs, see figure 2.16). So, if flow sessions (i,j) are backlogged at time (τ), the service for flow session (i) equal the service for flow session (j). And also, if the bandwidth equal 1, it will be divided equally for both sessions (session (i) = 0.5 and session (j) = 0.5). On the other hand, if there is one backlogged session, it will consume all the bandwidth.

Therefore, GPS can be described with all flows have equal weights as the following equations (2.7, 2.8, 2.9):

$$S(i,\tau,t) = S(j,\tau,t) \tag{2.7}$$

$$g(i) = \tfrac{R}{N} \tag{2.8}$$

$$r(i) \geqslant \tfrac{R}{N} \tag{2.9}$$

Where:

- $S(i,\tau,t)$: is the service of flow (i) in $[\tau,t]$, flow (i) is active at this interval.
- $g(i)$: is the minimum rate guaranteed for flow (i).
- R: is the total bandwidth.
- N: is the number of active sessions.

We will illustrate the GPS calculations and its packet emulation (PGPS) with the following example:

An example of GPS and PGPS, which describes the behavior of two sessions (flows). Every session belong to a buffer queue, see table 2.5. As described, the time of arrival for each packet arrives at the head of the queue. Each packet is served depending on the finish time for both GPS and PGPS.

Table 2.5: An example of GPS and its packet emulation PGPS that illustrates the comparison between two sessions.

	Session1				**Session2**		
Time of arrival	1	2	3	11	0	5	9
Size of Packet	1	1	2	2	3	2	2
Finish-time (GPS)	3	5	9	13	5	9	11
Finish-time (PGPS)	4	5	7	13	3	9	11

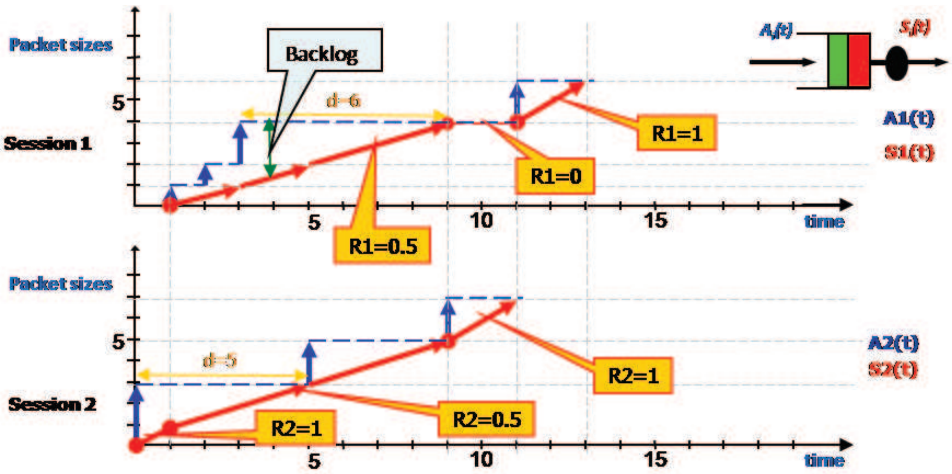


Figure 2.16: GPS serves two sessions.

As seen in figure 2.16, where:

- R1: average rate for session1.
- R2: average rate for session2.
- Si(t): GPS serving method for session i (S(i,t)).
- Ai(t): arriving packet for session i (A(i,t)).
- Backlog: there is waiting a packet.
- Packet sizes in units (u).
- Bandwidth = 1 u/s.
- The slope of the curve is equal to the average rate of session (r(i)):
  - r(i)= 1 u/s when the other session of i is idle.
  - r(i)= 0.5 u/s when there is no idle sessions.
  - r(i)= 0 u/s when the session of i is idle.

In figure 2.16, at time 0 session 2 serves the first packet. Since it is the only active session. Thus, it is served with all of the bandwidth (rate R2=1). Then, at time 1 session 1 serves the first packet, so the bandwidth is distributed half for session 1 and the second half for session 2, and so on.

In the other hand, figure 2.17, it presents the practical method (PGPS) which is the emulation of GPS. So, at time 0 session 2 serves the first packet with rate R2=1. Then, at time 1 session 1 backlogs the first packet, it cannot serve the packet until session 2 finish serving its packet at time 3, and so.

As illustrated above, GPS calculations are approximated by PGPS. However, there is difference for serving delay packet. For example, the delay for the third packet of session 1 in GPS is equal to 6 and in PGPS is equal to 4. And the difference is compensated for serving other packets.

However, it is possible to make worst-case network queuing delay guarantees constrained by leaky bucket traffic shaping. And also, it is independent

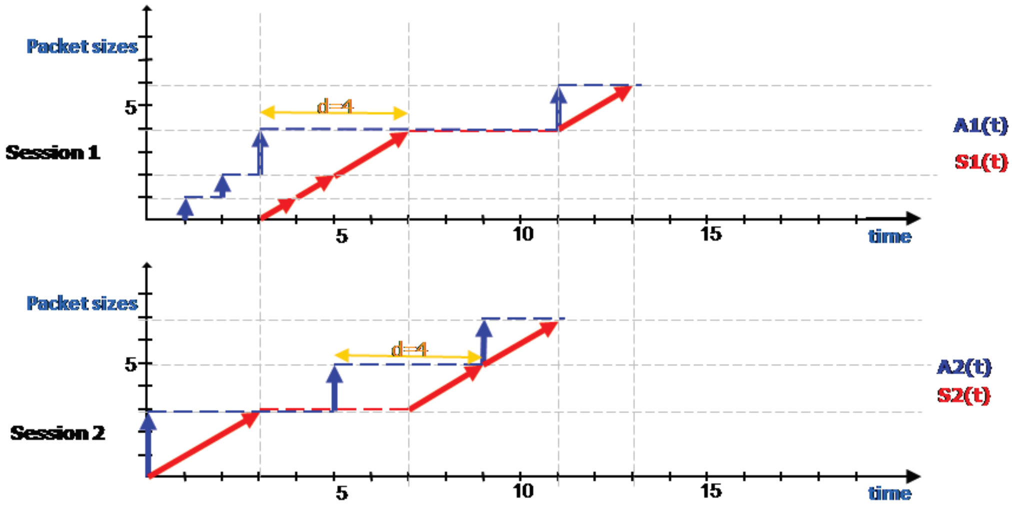


Figure 2.17: PGPS serves two sessions.

of the queues and the arrival of the other sessions. Therefore, every session is isolated and has its queue. So, GPS discipline can provide an end-to-end bounded delay service to a session whose traffic is constrained by a leaky bucket, because GPS is an ideal fluid model that serves infinitesimals. Practically, in PGPS, the worst-case of queuing delay is bounded by the maximum size of queued packet.

As a consequence, PGPS is an emulation for GPS and referred by many mechanisms (i.e. WFQ, VT, WF2Q and SCFQ). We will describe WFQ, because it is a simple mechanism that approximates and emulates GPS. Thus, its concepts would be adopted in our model (FQ-EDCA), see chapter 4.

2.7.7 Weighted Fair Queuing

WFQ is the ideal scheduler (Kesh 89), which is taken as a reference to compare other schedulers, is GPS (General Processing Sharing) (Pare 93). It ensures fair resource allocation for all backlogged sessions. And also, it uses an idealized fluid model that serves infinitesimals. That means, GPS is not implementable because real schedulers serve packets are not bits-based. Therefore, various packet approximation algorithms of GPS have been proposed. One of the best approximations is WFQ (Weighted Fair Queuing) (Kesh 89).

In fact, WFQ runs GPS computations and uses them to tag packets in flows to indicate the time that would have finished. WFQ approximates GPS by using byte-by-byte round-robin service. Thus, for a packet of size N bytes at the head of a queue, it will have complete service after the scheduler has performed N rounds of visiting all the other queues. Actually, packets cannot be transmitted one byte at a time from multiple queues. So, instead the packets are tagged with a finish-number, which indicates when they would have completed service if served byte-by-byte. Therefore, we can see that packets with the smallest finish number are the ones that should be transmitted first. Let us take the simple case that the queue is empty and a packet of size 100 bytes arrives in the queue. When it is served byte-by-byte round-robin, it will be serviced in 100 rounds. So, its finish number will simply be the current round-number plus its own size in bytes. Thus, if the round-number is 50, the finish number for that packet is 150. If the queue is non-empty, then the packets, already in the queue, will be served first. Thereby, the finish number of the next packet will be the finish number for that packet in front of it plus its own size in bytes. And also, when the round-number is equal to (or greater than) the finish number of a packet, it should be served immediately. However, the rate of round-number is controlled by the number of active connections.

Therefore, the finish number for packet k on flow i at time t is given by the following equations (2.10, 2.11):

$$F(i,k,t) = \max\{F(i,k-1,t), R(t)\} + P(i,k,t) \tag{2.10}$$

$$R(t) = R(\tau) + (t-\tau)/N_{ac}(t,\tau) \tag{2.11}$$

Where:

- F(i,k,t): is the finish number of packet k for flow i at time t.
- P(i,k,t): is the size of packet k for flow i at time t.

- R(t): is the round number at time t.
- R(τ): is the round number at time τ.
- N_{ac}(t,τ): is the number of active flows in the interval [t,τ].

In particular, the bounded delay provided by WFQ (PGPS) is within one packet transmission time of that provided by GPS. Even though it is computationally expensive, and need to hold each flow separately, WFQ is used by many router manufacturers.

However, there are alternatives to WFQ such as: self-clocked fair-queuing (SCFQ) (Gole 94), Virtual-Clock (VC) (Zhan 91) and Worst-case Fair Weighted Fair Queuing (WF2Q) (R 96). SCFQ has similar properties to WFQ, except that it removes the expensive round-number computation (by time-stamping synchronization). But also, it has a higher end-to-end delay bound than WFQ. Whereas, VC has a major difference of WFQ; that VC is based on the resource reservation instead of an equal share to all users. That can allocate any specific resource to each user. Whenever, WF2Q has more approximation fluid flow of GPS. That it provides almost identical service with GPS. And also, the maximum difference between GPS and WF2Q is no more than one packet size, because WF2Q serves all the backlogs sessions.

In other words, in a WFQ system, when the scheduler chooses the next packet for transmission at time t, it selects among all the packets that are backlogged at t, and selects the first packet that would complete service in the corresponding GPS.

In the other hand, in a WF2Q system, when the scheduler chooses the next packet at time t, it chooses only from the packets that have started receiving service in the corresponding GPS at t, and chooses the packet among them that would complete service first in the corresponding GPS. Thus it has less end-to-end delay bound than WFQ.

However, in the example above, which illustrates the GPS and PGPS. As seen in figure 2.17, the behavior of PGPS can simulate both of WFQ and WF2Q

As illustrated above, WFQ and WF2Q have almost the same properties. But, WFQ has a problem, and WF2Q has better approximation of GPS than WFQ. The problem of WFQ is that it serves the backlogged packets. So, if there are condensed backlogged packets in a queue, it will serve these packets and the other non-empty queues will wait serving. Therefore, instead of searching for the smallest finishing time of all packets in the queue, WF2Q looks for the packet with the smallest finishing time, of whose start time has already occurred.

As described in chapter 4, the principles of calculations of WFQ and WF2Q are adopted in FQ-EDCA and transfered in WLAN ad-hoc to apply fairness between all source nodes.

2.8 Conclusion

To support true end-to-end QoS on an IP-network, two technologies have been defined: Integrated Services (IntServ) and Differentiated Services (DiffServ). IntServ follows the signaled QoS technology, where the end-hosts signal their QoS needs to the network. While DiffServ works on the provisioned QoS technology, where network elements are setup to service multiple classes of traffic with varying QoS requirements. Thus, these technologies need to be translated in wireless networks by adopting their principles such as differentiating, marking, classifying, shaping, scheduling, etc.

However, wireless scheduling plays an important role in the design of wireless networks protocols. As well, it determines the QoS provisioning over the wireless link. So, it is very clear that the scheduling algorithm is the core of the protocol that determines the performance and guarantees QoS. However, the design of an ideal wireless scheduling is assumed by a full knowledge on current channel conditions and analytically provable delay and throughput bounds. Therefore, it is not simple to adapt a model that is designed for wired environments. And also, it is assumed, that the wired networks have error-free channel. While, the wireless networks have burst error channel and location dependent channel. Thus, a lot of research is oriented on extending the scheduling algorithms proposed in wired networks to wireless domain as reported in (Lu 97).

Thereby, to develop a model for wireless networks, it must supply; a fairness model for wireless fair queuing, knowledge on current channel conditions and approximates the characteristics of the ideal algorithm.

However, it is pronounced that the MAC layer needs substantial more work to have a complete solution for wireless fair queuing. Typically, most related works of fairness in wireless networks concentrated on the probability of channel access and controlling the contention window. While in wired networks, fair scheduling has been extensively researched. So, fairness schedulers should be affected carefully with the probability of channel access and contention window.

As a conclusion, the scheduler mechanisms (Simo 03), that provide QoS have to support services for arrival packets as: defining and keeping separate queues for different flows, sharing bandwidth for resource allocation and bandwidth guarantees, reducing delay variations (Bounding Delay) to guarantees QoS, *isolating ill-behaved users and applying the fairness to competing flows.* We will see that FQ-EDCA (Chapter 4) takes into account the previous services, notably the last crucial service. In the other hand, for outgoing packets, they manage serving packets in the output queues. Thus, the QoS could be guaranteed, and the congestion of the network could be avoided too (Floy 93; Kesh 92).

However, an extension is needed to enhance RAMAC behavior when the nodes in the network use other different traffic types with different priorities, which can be implemented to guarantee QoS with fairness of bandwidth sharing. Therefore, our model (F-EDCA) evaluates and extends RAMAC, and enhances QoS (EDCA) in wireless ad-hoc networks, as we will see in chapter 3. While, FQ-EDCA applies fairness and enhances the QoS by implementing a hierarchical scheduling technique, as described in chapter 4.

Further, both of F-EDCA and FQ-EDCA are enhancements for MAC layer. While, routing protocols are applied over MAC layer to satisfy an end-to-end QoS requirement for mobile hosts wireless networks (MANET). However, the mobility of host node make the topology of ad-hoc network is dynamically varying. That makes frequent paths breaks and discovers another paths. This re-establishment of paths may be unacceptable for some of application because of extra delay, packet losses, etc. For that reason, our enhancement of F-EDCA and FQ-EDCA are applied and performed on static topologies to avoid the interruptions of routing paths breaks. This could evaluate a pure study of fairness resource allocations for EDCA in MAC layer

Part II

New Models for Enhancing QoS with Fairness

Chapter 3

F-EDCA

F-EDCA (Abuz 09c; Abuz 08) is a fairness model for QoS optimization in wireless network. It aims to develop and optimize the QoS with fairness in wireless ad-hoc networks. As discussed in Chapter 2, IEEE802.11e (EDCA) (IEEE 05a) enhances the original standard by enhancing the Distributed Coordination Function (DCF). Where, DCF consists in deferring the channel access in ad-hoc mode. Thereby, EDCA differentiates and classifies the packets in Access Categories depending on their priorities. Although EDCA is concerned to be the important enhancement for QoS in wireless ad-hoc networks, EDCA is not enough to provide fairness in QoS guarantees.

There are many propositions that develop the QoS in ad-hoc mode. They are based on control access channel such as Multiple Access Collision Avoidance with Contention Window optimization (MACAW) (Bhar 94). This model is enhanced with (RAMAC) (Nait 03) to access channel fairly by differentiating the routed and owned packets in routing nodes. However, RAMAC is implemented in the legacy DCF to obtain fairness. It does not take into account the EDCA model. Thereby, we had proposed F-EDCA to enhance and develop QoS of EDCA in ad-hoc networks by implementing the principles of RAMAC. Thus, the transmitted queued packets in routing nodes will be differentiated depending on their priorities (EDCA specifications) and their type (Owned or Routed packet) to apply fairness in routing nodes.

As a consequence of the spread of popularity in WLANs, the interest for ad-hoc networks is taken into account. Where, the wireless devices need to transmit their packets with fairness and guaranteed services (i.e. Multimedia packets) in ad-hoc networks.

In addition, ad-hoc uses peer-to-peer through single-hop or multi-hop paths. Therefore, each station must work cooperatively to apply the connectivity between the sender and the receiver. Thus, each station must accept to route and forward the packets of others stations.

Although this is the only solution in ad-hoc to apply the connectivity, many of cases this leads to unfair resource allocation. Because some of stations (Routing nodes) contend to access channel for transferring their packets (Owned Packets) and the packets of their neighbors (Routed Packets) to apply a complete cooperation. Other stations (Non-routing nodes) contend to access channel for transmitting only their own packets. Indeed, there are nodes that have more bandwidth than others. As a consequence, unfairness of resources allocation minimize QoS guarantees notably for routing nodes.

Indeed, there is no admission control that manages all traffic for all hosts in ad-hoc. Every host station manages its data separately, which makes QoS in ad-hoc is a hard and complicated challenge. Therefore, in ad-hoc, it cannot guarantee QoS for all conditions and for all the time. Thus, we need to address each condition separately to enhance QoS guarantee in ad-hoc.

Moreover, the EDCA does not have a distributed admission control algorithm to access the channel fairly. Therefore, we are working on an enhancement of the EDCA, such that it can be used to provide fairness of resources allocation for routing nodes. Thus, every node can differentiate between its owned packets and the routed packets, which are queued depending on packet's priority. As well, we implement F-EDCA taking the advantage of RAMAC concepts in EDCA. Thereby, by adding this service, routing nodes can access channel more frequently than the other nodes. As a result, F-EDCA supports fairness of resource allocations in ad-hoc wireless networks.

Thus, QoS guarantees will be optimized in EDCA.

In this chapter, we will first describe our model (F-EDCA) that enhances EDCA with regard to fairness. The next section gives simulations results for the two models: F-EDCA and EDCA with its default parameters. These comparative simulations of the two models will enable us to evaluate the real benefits of F-EDCA. These simulations will be performed under conditions which allow comparing their performances and their capacity to guarantee fairness.

Most of the proposals for QoS enhancement in ad-hoc wireless networks mainly focus on two conceptions; the station-based DCF enhancement scheme such as (Nait 03; Bhar 94; Aad 01) and queue-based enhancement scheme such as (IEEE 05a; Mang 02; Romd 03; Wiet; Abuz 09a). Thereby, we will present F-EDCA, which is a fusion of the last two conceptions.

The principles of DiffServ for differentiating flows is adapted and translated in EDCA by introducing the concepts of Access Category (AC). Thereby, we will adopt the conception of RAMAC in EDCA. This adoption is not simple, because there is no compatibility between the parameters of EDCA and RAMAC. Thus, F-EDCA will take into account this compatibility. As a consequence some of ideas in F-EDCA will try and describe this compatibility, as presented in the next section.

And also, an extension is needed to enhance RAMAC behavior when the nodes in the network use different traffic priorities to transfer the data of different services. It leads us to implement this idea in EDCA in order to guarantee QoS with fairness resource allocations. Therefore, we will present F-EDCA (Fairness EDCA) and its evaluations that extend RAMAC and enhance EDCA in wireless ad-hoc networks.

3.1 Definition of F-EDCA

First, we should define some of the main notions used by F-EDCA. There are two types of nodes in WLAN ad-hoc that are:

- Routing node: it is a wireless node station that routes the packets' flows of its neighbors. The packets in this node are classified, queued depending on their respective class and differentiated to:
  - Owned packet (O): It is a packet that belongs to this node.
  - Routed packet (R): It is a packet that belongs to a nearby node that is routed by this node.
- Non-Routing node: It is a wireless node station that does not forward its neighbors packets.

Let us precise here that in fact in an ad-hoc network, most of the nodes are routing nodes. At the borderline there are few nodes that are just source non-routing nodes.

However, F-EDCA means Fairness EDCA. Thus, a fairness model is applied in EDCA to enhance it. Further, the Fairness is one of the topics that can enhance the QoS guarantees. The fairness objective in wireless network depends on its topology, its size, its technology and the type of traffics and their applications. Therefore, it must follow an agreement (i.e. SLA). However, the fairness objective could be referred to many considerations as described in the general introduction, such as flow fairness, prioritized flow fairness, source fairness, etc.

Thus, the bandwidth is considered to be fairly allocated as described in (Gamb 04). As well, EDCA provides packet level fairness (Priority classes) in MAC to guarantee QoS. While there is a lot of research work on improving fairness in the presence of hidden terminals (Bhar 94; Nait 03; Toba 75). So, the fairness objective in F-EDCA is the source fairness. It is defined as the bandwidth or throughput of each source node that must be allocated fairly between all the sources. Thus, in F-EDCA, the resource is fairly allocated with respect to the percentage of owned packets and the priority of their respective class.

3.2 F-EDCA model

QoS architecture of EDCA must be provided in WLAN to make F-EDCA applicable. Therefore, we will extend to EDCA the concepts of fairness that is developed by RAMAC.

Thereby, we will enhance EDCA to get more fairness for routing nodes. This can be realized by adapting a mechanism that controls the routing nodes to access channel more frequently than non-routing nodes. So, the allocated bandwidth of routing nodes, which is consumed for forwarding packets, will be compensated. For example, figure 3.2 presents the distribution of the bandwidth between three nodes (A, B and C). A and C are non-routing nodes, and B is a routing node, for the topology as seen in figure 3.1. In figure 3.2.a, the bandwidth is distributed equally between A, B and C. Indeed, the allocated bandwidth of B for transmitting its owned packets is less than the allocated bandwidth of A or C. This can be adjusted, see figure 3.2.b, by allowing more bandwidth allocated to B. Therefore, A, B and C get nearly the same allocated bandwidth to transmit their owned packets. Thus, the obtained fairness can minimize the delay and increase the QoS in ad-hoc networks.

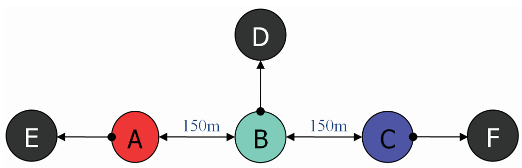


Figure 3.1: The topology of simulations (node (B) is a routing node, nodes (A,C) are non-routing nodes and nodes (D,E and F) are sink nodes)

In wired networks, unfairness of bandwidth allocation problem can be solved simply by enhancing the scheduling techniques (i.e. Weighted Fair Queuing (Kesh 89), Weighted Round-Robin (Kate 91), . . .). As explained, it is a hard challenge to adapt wired techniques in wireless. In other words, a lot of proposals suggest another solution to manage bandwidth allocation by accessing channel. Thus, the contention window (CW) will be controlled

Figure 3.2: BW allocation for routing node (B) and non-routing nodes(A,C)

in F-EDCA based on a MACAW/RAMAC like approaches as described in (3.1). However, F-EDCA defines three parameters for routing nodes: a ratio of fairness (ρ) as described in (3.2), a Multiplicative Factor (MF) and a tuning constant parameter (β) as described in (3.3). As well, all the following equations are applied in each access category (AC) for node i:

$$CW_{new}(AC,i) = min(MF^e(AC,i) * CW_{old}(AC,i), CW_{max}(AC,i)) \quad (3.1)$$

$$\rho^e(AC,i,t) = Wo(AC,i,t)/(Wo(AC,i,t) + Wr(AC,i,t)) \quad (3.2)$$

$$MF^e(AC,i) = 1 + \beta(AC) * \rho^e(AC,i) \quad (3.3)$$

Where:

- $MF^e(AC,i)$ is the enhanced multiplicative factor for AC in node i.
- $\beta(AC)$ is the constant tuning factor for AC in each node.
- $\rho^e(AC,i,t)$ is the enhanced ratio of the owned packets over the total queued packets in AC for node i.
- $Wo(AC,i,t)$ is the amount of owned packets to send in bytes, which belong to node i and queued in its AC at time t.

- $Wr(AC, i, t)$ is the amount of routed packets to send in bytes, which belong to other nodes and queued in AC of node i at time t.
- (t) is the time of unsuccessful access channel (collision).

Indeed, equations (2.4, 3.3) have the same goal, which is to make the routing node access the channel more frequent than non-routing node. However, in F-EDCA, we have modified the equation (2.4) to (3.3), because MACAW, RAMAC and other approaches are proposed and implemented for pure DCF. These approaches modified the parameters of CW_{min} and CW_{max}. For example, RAMAC had modified the parameters of the legacy DCF by reducing CW_{min} from the value 31 to 3. This change makes the current CW exceed the CW_{max} slowly after consequent collisions. Thereby, this makes the parameters of equation (2.4) be significant. Thus, if the current CWs of the collided nodes (i.e. routing and non-routing nodes) exceed the CW_{max}, they will be equal to CW_{max} for both of the collided nodes. Therefore, these nodes will have the same possibility with no priority for routing nodes to access the channel.

On the other hand, EDCA classifies the packets to AC depending on their priority. Whereas, the parameters of CW_{min} and CW_{max} for the highest priority are equal to 7 and 15 respectively, as described in table 2.3. So, after two consequent collisions between nodes (i.e. routing and non-routing nodes), the current CW of these nodes will exceed the CW_{max}. This makes the parameters of the equation (2.4) are insignificant and makes MF almost be doubled with no priority for routing nodes. This leads to access the channel with the same chance for both routing and non-routing nodes, almost like the legacy EDCA. As a conclusion, there are two ideas to solve partially the lacuna described above. Firstly, by modifying the equation (2.4) to (3.3) to make the value of MF near to be 1 for routing nodes. Secondly, modifying the parameters of CW_{min} and CW_{max} for each AC to allow the parameters of equation (3.3) be significant. So, after modifying the initial definition of MF, it still has the fairness assured by (2.4), as illustrated in section 3.5.

However, the tuning parameter ($\beta(AC)$), noted here for simplicity, is a constant parameter for all the nodes that depend on the priority of class (AC). Its value is ranged between 0 and 1 (to define the value of $MF^e(AC, i)$ between 1 and 2). If the value of β_{AC} near to 1, the value of $MF^e(AC, i)$ will be maximized. As well, if the value of β_{AC} near to 0, the value of $MF^e(AC, i)$ will be minimized. Thereby, β_{AC} must be minimized for the high priority class depending on the parameters of CW_{min} and CW_{max}. However, the value of β_{AC} is regulated depending on the AC's parameters of EDCA (i.e. CW_{min},CW_{max} and AIFSN) performed by the simulations. So, regulating β_{AC} with modifying CW_{min} and CW_{max} for all ACs is a hard work and need a lot of time.

For example, β_{AC} is minimized for the high priority class and equal to 0.5. This value is significant when the parameters of CW_{min} is equal to 3 and CW_{max} is equal to 15. These values of parameters are performed by the simulations, as described in section 5.1, and achieve the required fairness as the result obtained in section 3.5.

As well, for the sake of simplicity, β_{AC} is treated as a fixed number for each AC evaluated by the simulation. Thereby, F-EDCA can assure the required fairness for each AC. Let us recall here that the default RAMAC, as described in equations (2.3, 2.4, 2.5), cannot be adapted to apply fairness for each AC with respect to EDCA parameters because it does not do differentiation between data and so does not consider their different priorities. So, F-EDCA make the compatibility between RAMAC and EDCA by equations (3.1, 3.2, 3.3)

3.3 The working principles of F-EDCA

As seen in figure 3.3, each routing node station apply the following algorithm in MAC layer:

- It differentiates between its owned packets and the routed packets by marking with (O) and (R) respectively. An interface between upper

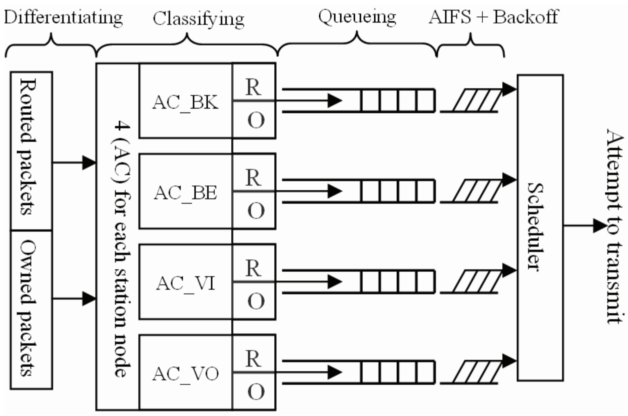


Figure 3.3: Model of F-EDCA

layers (i.e. Network layer) and MAC layer is supposed to differentiate the packets in the different access classes. But this problem is out of the scope of this paper.

- Each packet is classified and queued according to its priority class.
- The value of ρ is calculated by (3.2) for each AC when there is a collision. This is done with taking into account the size and the type of packets queued in each AC. Therefore, CWnew (3.1) is obtained.
- Each packet at the head of queue waits its back-off and AIFS.
- If several packets at the head of the different queue have the same waiting time before its access to the medium, a virtual collision will occur inside the node. Consequently, a simple scheduler (like Round-robin) treats this problem. Then the node attempts to transmit the head packet of the selected queue.
- CW is reinitialized to its minimum value after each successful transmission.

3.4 The simulations of F-EDCA

Network Simulator 2 (version 2.31 of ns-allinone (NS2)) is chosen to implement and evaluate F-EDCA. The topology, which is used in our simulations, is described in figure 3.1. This topology is a basic ad-hoc network. Its simplicity allows a better understanding of the bandwidth fairness allocation between the routing node and the others. So we can treat it like a generic case in ad-hoc wireless network area that focuses the problem of unfairness of bandwidth allocation. Indeed, this is the objective of the simulations, because without dealing with this problem, F-EDCA would work like EDCA. And also, another objective, the difference between routing and non-routing nodes could be differentiated by these simulations.

However, Nodes (A, B, C, D, E and F) are nodes working in ad-hoc mode. B is a routing node. A,C, D, E and F are non-routing nodes. D, E and F are treated as sink nodes. E and F are implied to receive flows of A and C respectively. This leads to avoid the impact of hidden terminal problem that is described in (Kose 08). Indeed, the hidden terminal nodes may need to transmit multiple flows to their neighbors (one-hop destinations) or to be forwarded through routing nodes to multi-hop destinations. So, if A, B and C send to the same destination (ex. F), another problem will be produced, as described in section 4.1. Thereby, the topology of figure 3.1 is chosen to separate the other problems, and to focus on the problem of unfairness between routing and non-routing nodes.

Our aim in these simulations is to compare the respective bandwidth of nodes A, B and C. Six flows are sent by these source nodes (A, B and C) as the following:

- A sends a flow to C routed by B. And another flow for its neighbor E without routing by B.
- B sends two flows to its neighbor D.
- C sends a flow to A routed by B. And another flow for its neighbor F without routing by B.

Table 3.1: The parameters of simulations

Standards	IEEE802.11b/e
QoS	EDCA
Topology mode	ad-hoc
Routing protocol	DSDV
Routing nodes	B
Non-routing nodes	A and C
Sink nodes	D, E and F
Number of flows	6 (All flows have the same priority)
Type of flows	CBR/UDP
Distance between nodes	150m
Transmission range	250m
CWmin	3
CWmax	1023
β	1
Packet size	1472 bytes
Packet interval	[5ms, 10ms]
Bandwidth	11Mbps
Time of simulation	200s

The flows are sent at constant bit rate CBR/UDP traffic applications with constant packet size (see table 4.9). And also, each source node sends two flows to have the same conditions for accessing the channel. In addition, F-EDCA is compared with EDCA model default parameters (Wiet). To achieve the comparison, bottlenecks and heavy loads of traffic flow should be applied in the network. This is the case that is considered in the performed topology. All the parameters of simulations are presented in table 4.9. The values of the parameters are chosen to focus clearly the problem of unfairness between routing and non-routing nodes and to be treated by F-EDCA.

3.5 The results of first simulations

The results of our simulations are considered as following:

Figure 3.4 describes the bandwidth allocations for default EDCA model (Wiet). We remark that although the allocated bandwidth to transmit all the packets (owned and routed) queued in B is nearly as much as A and C, the dedicated bandwidth to transmit the owned packets of B is weak compare with other source nodes (i.e. A and C). Indeed, all source nodes (A, B and C) obtain nearly 2.4Mbps of bandwidth.

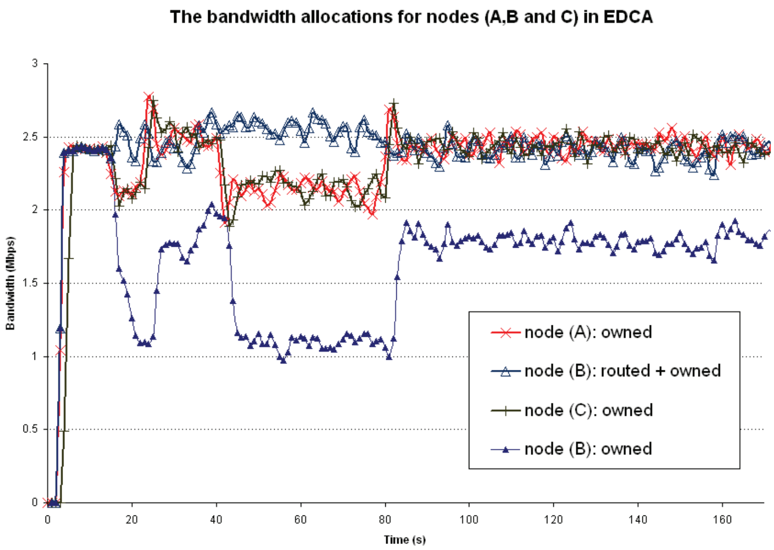


Figure 3.4: EDCA default behavior (without fairness)

Nevertheless the allocated bandwidth for the owned packets of B is nearly 1.8Mbps. Therefore, there is unfairness of bandwidth allocations.

Figure 3.5 describes the bandwidth allocations by using F-EDCA. We remark that the allocated bandwidth of B is bigger than those of A and C. But the allocated bandwidth of the owned packets for each node is the same. Indeed, every source node gets nearly 2.3Mbps of bandwidth for its owned packets.

As well, we observe that the allocated bandwidth for all the packets of B is increased nearly 40% (from nearly $2.4Mbps$ to around $3.3Mbps$). While the allocated bandwidth of A and C is slightly reduced.

However, when B contend to access the channel more frequently than others, the collisions is increased. Thus, figure 3.5 shows more variations than figure 3.4.

Figure 3.6 explains the described phenomenon previously. Indeed, when the allocated bandwidth of B is increased, the dedicated bandwidth for the

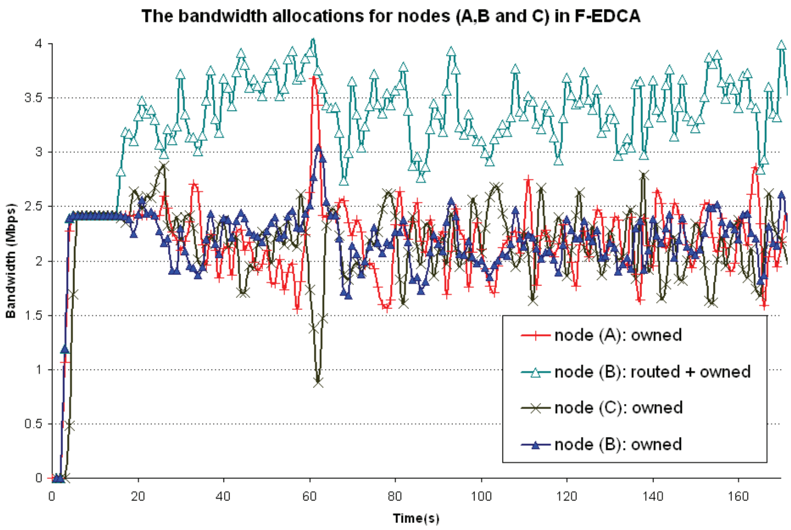


Figure 3.5: F-EDCA behavior (with fairness)

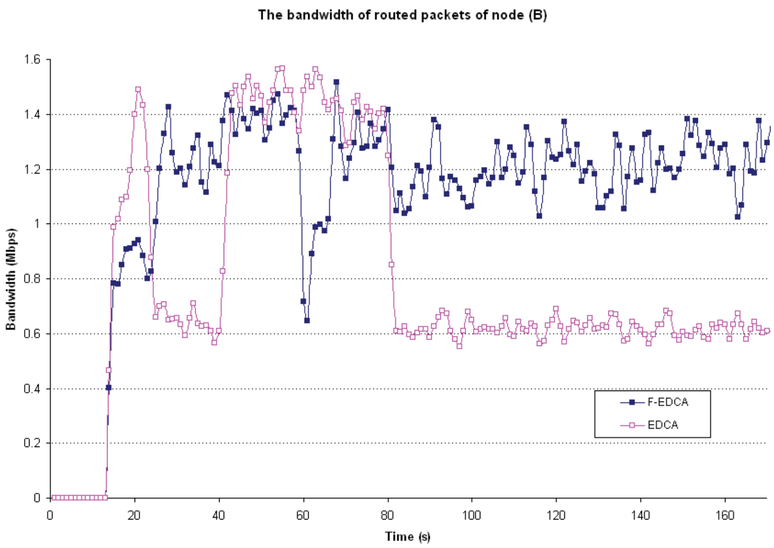


Figure 3.6: Throughput of routed packets in node (B)

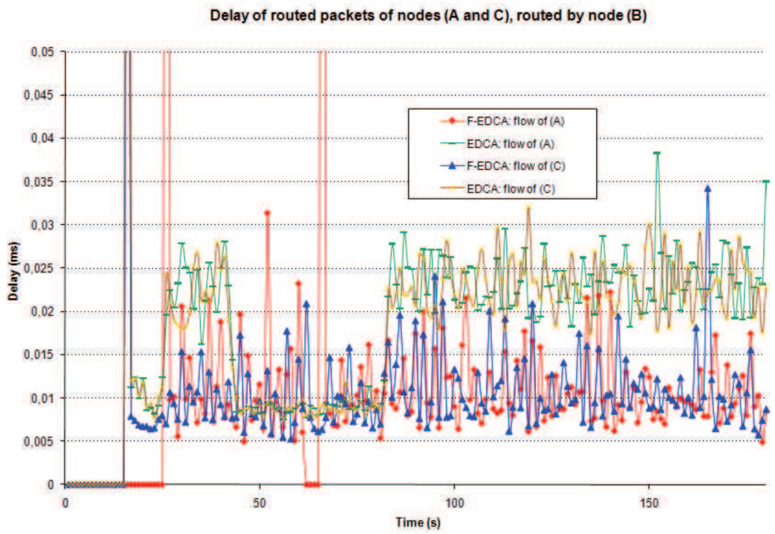


Figure 3.7: Delay of flows routed by node (B)

routed packets, which belong to A and C, is increased also. Therefore, the allocated bandwidths of A and C are slightly minimized.

Figure 3.7 describes the delay of routed packets of B. As well, figure 3.8 describes the delay of owned packets of B for both F-EDCA and EDCA. It is observed that the delay is decreased by 50% for the two types of packets in F-EDCA compared with EDCA. Thereby, the QoS is optimized with F-EDCA.

Further, as seen in all figures, that F-EDCA achieves the stability of resource allocations better than EDCA. So, F-EDCA became stable after almost 20s. This time is needed to ad-hoc networks for constructing the routing paths. While, EDCA became stable after 80s due to constructing the routing paths, the congestion, hidden terminal problem, etc. Therefore, we can say that F-EDCA treats partially these problems

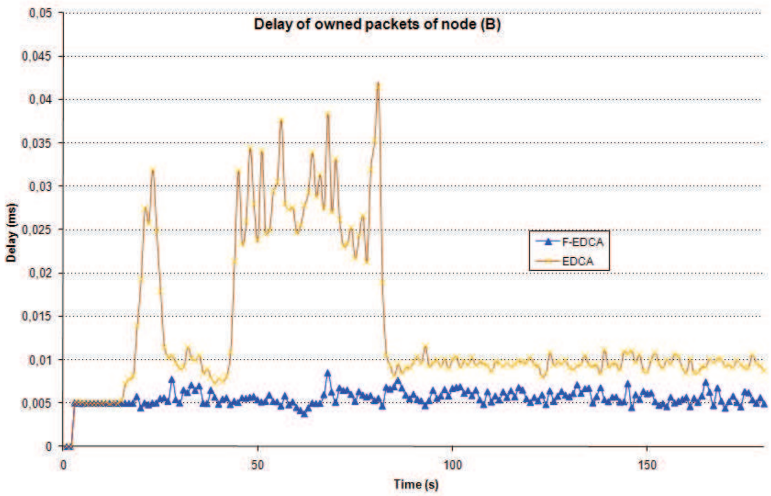


Figure 3.8: Delay of node (B)'s flow

3.6 Conclusion

Many of researches of QoS enhancement in ad-hoc wireless networks mainly focus on station-based DCF enhancement scheme or queue-based enhancement scheme. In our proposition (F-EDCA), we have merged the station-based DCF and queue-based enhancements. Therefore, we have implemented F-EDCA that supports bandwidth allocation with fairness in the architecture of EDCA by taking into account differentiation between wireless node stations (Routing and Non-routing nodes). Thus, the routing node can access the channel more frequently than other nodes when there is a collision. As a consequence, each routing node has an extra bandwidth (depending on the percentage of routed packets) to compensate the deduction of its bandwidth that is needed for forwarding neighbors' packets.

Therefore, we can say that F-EDCA treats partially the hidden terminal problem and avoids the congestion by allocating fairly the resources. Thus, the misbehaving of resource allocations can be resolved by F-EDCA.

Some perspectives of this study is to evaluate, to verify the scalability of this model and to propose a model to be able to guarantee end-to-end QoS in ad-hoc networks, as presented in chapter 4. And also, this can be done by regulating β with modifying the parameters of CW_{min},CW_{max} and AIFSN for each AC.

Chapter 4

FQ-EDCA

FQ-EDCA is a Fair Queuing Model for EDCA to optimize QoS in Ad-hoc Wireless Network. It aims to develop and optimize the Quality of Service (QoS) with fairness in multi-hop wireless ad-hoc networks. As we described in the state of the art, the development of the wireless networks is crucial. And the interest for ad-hoc networks has also increased. Because the needs for wireless devices (i.e. PDAs, Sensors, Wireless stations, ...) is rapidly increased, while their prices are decreased. Therefore, some of applications need to transfer critical data (i.e. Multimedia, real-time, ...) by using these devices. These data sometimes require strict services, which could not be offered by the legacy standard IEEE802.11 (IEEE 97).

Since, QoS supports critical and time-sensitive data, it is one of the challenges, that must be overcome, to realize the practical benefits of wireless networks specially ad-hoc.

As a consequence, an enhancement of that standard (IEEE802.11e) is developed and concerned to support QoS in both the infrastructure and ad-hoc modes (IEEE 05a). Therefore, for ad-hoc, IEEE 802.11e standard enhances the legacy IEEE 802.11 MAC layer to support QoS by introducing the Enhanced Distributed Channel Access (EDCA) that manages QoS in each station. EDCA is implemented and applied in ad-hoc to support services for multimedia traffic and real-time. It differentiates and classifies the packets in Access Categories (AC) depending on their priorities.

However, some of problems could be appeared in EDCA and not solved in F-EDCA, as defined in the next section. Therefore, FQ-EDCA is proposed to solve these problems.

In this chapter, we will present definitions and problematic in wireless networks, FQ-EDCA model and its algorithm. This model will be simulated to obtain the results and applied for many topologies. Therefore, these results will be evaluated and compared with the EDCA.

4.1 Definitions and problematic

The QoS could be satisfied by the standard IEEE802.11e, specially in the infrastructure mode. Although, in ad-hoc, the distributed EDCA is an important enhancement for the legacy IEEE 802.11, it is not enough to provide strict QoS guarantees required by real-time and multimedia services. Because there are crucial problems (i.e. hidden terminal) that need solutions, as explained in chapter 1, see figure 1.7, and in (Zhai 05).

In addition, another problem could appear frequently that an ill-behaved transmission of a source node leads to a lack of resource distribution and unfairness, because it could consume the majority of bandwidth and affect the performance of many others.

This problem could be described by this example, as seen in figure 4.1. Nodes A and B are source nodes that need to transmit their packets to node C. Thus, node B is defined as an in-ranged node, because it is in the range of transmission of node C. So, it can send its packets to node C directly. While, node A is defined as an out-ranged node, because it is out of the range of transmission of node C. So, it could not send its packets to node C directly and needs node B to route them. Like this situation, node B will attempt to send its packets and the packets of node A. That leads to the resource allocation of node A is ill-behaved controlled by node B while the congestion. So, node B is defined as an ill-behaved source node.

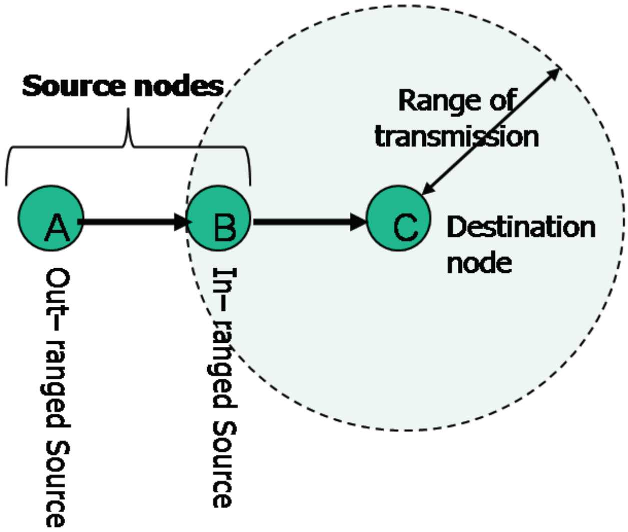


Figure 4.1: An example to define the ill-behaved sources

Generally, an ill-behaved source node is a source node that forwards its packets and routes the packets of its neighbors to the next-hop node or destination. Therefore, an in-ranged source node covered by its destination (one-hop source node), could be an ill-behaved source node, if it is routing the packets of an out-ranged source node uncovered by the same destination (multi-hop source node).

As a result, while bottleneck or state of starvation occurring, if the traffic is properly regulated in IEEE802.11e, it will be capable for supporting the QoS requirements for the real-time and multimedia traffic. Therefore, IEEE802.11e needs proper network control mechanisms, as explained in (Chen 06).

Moreover, the EDCA does not have any distributed admission control algorithm to access the channel fairly. We are working on an enhancement of the EDCA by implementing a Fairness Queuing model (FQ-EDCA). Thus, it can be used to provide fairness of resources allocations for source nodes by regulating their traffic and controlling their ill-behaved transmission. Thus, every node, in each AC, differentiates the packets (by its type and its source node). These packets are queued depending on their source nodes. Then, a hierarchical scheduling algorithm is implemented to serve and control the packets transmission.

As a result, FQ-EDCA supports fairness of resources allocation between source nodes in ad-hoc wireless networks. It is proved and analyzed by experimental simulations for many topologies like (star, chain, tree, ...), see figure 4.4. Thus, the bandwidth is distributed fairly between all the source nodes. As a consequence, these nodes can access the channel fairly to transmit their flows with respect to EDCA specifications of Access Categories (AC) priorities. And also, the delay of flows of the same priorities is minimized and nearly fixed for all the source nodes. This leads to minimize the Jitter also. Therefore, the services guarantees of QoS are enhanced.

4.2 FQ-EDCA model

FQ-EDCA is a Fairness Queuing technique applied in EDCA to guarantee QoS. The fairness objective in wireless must consider the resource which is to be fairly allocated, as described in (Gamb 04). In other words, the source nodes that have to transmit their loads (the same load for each node) have the same bandwidth of resource allocation.

However, EDCA is applied in MAC layer which coordinates the efficient use of the limited shared wireless resource. But, the problems of hidden terminal still exist and make EDCA unfairness for some cases because of the ill-behaved source nodes. This is obviously illustrated in (Kose 08) and our simulations.

In (Zhai 05), many proposals are proposed to solve the last problems, but this leads to new problems such as wasting bandwidth and channel utilization (spatial reuse). Therefore, FQ-EDCA has the ability to overcome these problems, without regenerating undesired problems, and prevent the ill-behaved sources from arbitrarily increasing their rate.

In (Abuz 08) we have proposed F-EDCA (Fairness EDCA) to obtain a fairness solution by differentiating between routing nodes and non-routing nodes. F-EDCA is a fusion of station-based DCF enhancement with EDCA that leads to a fairness approach, while FQ-EDCA is a pure queue-based enhancement. Thus, the fairness solution in FQ-EDCA is obtained by differentiating the packets type and source nodes and by implementing an adaptive fair queuing scheduler. These implementations are applied in each AC separately to reduce the complexity of our algorithm.

4.2.1 The algorithm of FQ-EDCA

In FQ-EDCA model, hierarchical scheduling techniques (multi-level of scheduling packets) are implemented, to control network traffic and to guarantee QoS with fairness. The network traffic could be controlled by traffic rate control and traffic-shaping to optimize QoS guarantee performance, low latency and bandwidth. Therefrom, controlled traffic rate is performed with scheduling and queues disciplines.

In wired networks, traffic-shaping schemes are commonly implemented at the network edges to control traffic that is entering the network (DiffServ)(RFC2a; Lee 06). Which make the network has a proper admission control. While, in ad-hoc IEEE 802.11(e), every node is autonomous and there is no proper admission control, as described in (Chen 06; Zhai 05). Thus, to transfer the conceptions of wired networks DiffServ to ad-hoc wireless networks, we would implement a DiffServ like in each node station, as explained in FQ-EDCA algorithm.

It is very important for bounding the end-to-end delay to guarantee QoS requirements. Therefore, traffic-shaping provides a limited and controlled rate when queuing and serving the packets with a proper scheduler discipline.

The scheduler disciplines, such as GPS (Pare 93), WFQ (Kesh 89), WRR (Kate 91), WF2Q (R 96), etc., control the order for forwarding the packets. They can be implemented to avoid congestion and to distribute the bandwidth equally (with Fairness). However, these disciplines are successfully implemented and evaluated in wired networks. Therefrom, an extension of these disciplines is needed to perform its fairness in wireless network. Thus, their conceptions of fairness is translated and adapted to apply the expected fairness for EDCA in each AC. That we will see in FQ-EDCA algorithm.

The simple solution for scheduling is that the packet which first comes is first served (FCFS). It is used for best effort service (without QoS).

In the other hand the scheduler that provides QoS guarantee is a hard challenge, because it depends on the type of network and its topology, the protocols of upper layers, the type of flow, etc. Therefore, the model have to support services for arrival packets as: defining and keeping separate queues for different flows, allocating resource required for each flow, reducing delay variations (bounding delay), *isolating ill-behaved users and applying the fairness to competing flows*. We will see that FQ-EDCA takes into account the previous services, notably the last crucial service.

However, when EDCA receives packets from the network layer, it classifies the packets depending on their priorities that are assigned in upper layers. After the classification, in EDCA, the packets for the same AC are enqueued in First Come First Serve (FCFS) with drop tail technique (drop the last packet when the queue is full). While, in FQ-EDCA, multiple levels of scheduling are implemented to compensate EDCA by controlling traffic and apply fairness, see figure 4.2.

Then, for both FQ-EDCA and EDCA, the packet is dequeued when it can access the channel to attempt for transmission after a waiting time. This waiting time is the summation of deferring time of the Arbitrary Inter-Frame Number, depending on its AC (AIFSN(AC)), and the back-off countdown time. The back-off time is chosen randomly between the interval (CWmin, CWmax). These packets will be transmitted depending on the parameters of EDCA (AIFSN, CWmin and CWmax) (IEEE 05a). These parameters are decided depending on the policy for the QoS in wireless, which is already asserted. Thus, FQ-EDCA has the same parameters of EDCA to respect that policy.

Therefore, the preferred fairness must be applied in each class (AC) for each node. And also, a traffic rate control discipline must be performed. So, a fairness scheduling technique is developed with a separate queue for each source node in each AC, as seen in figure 4.2.

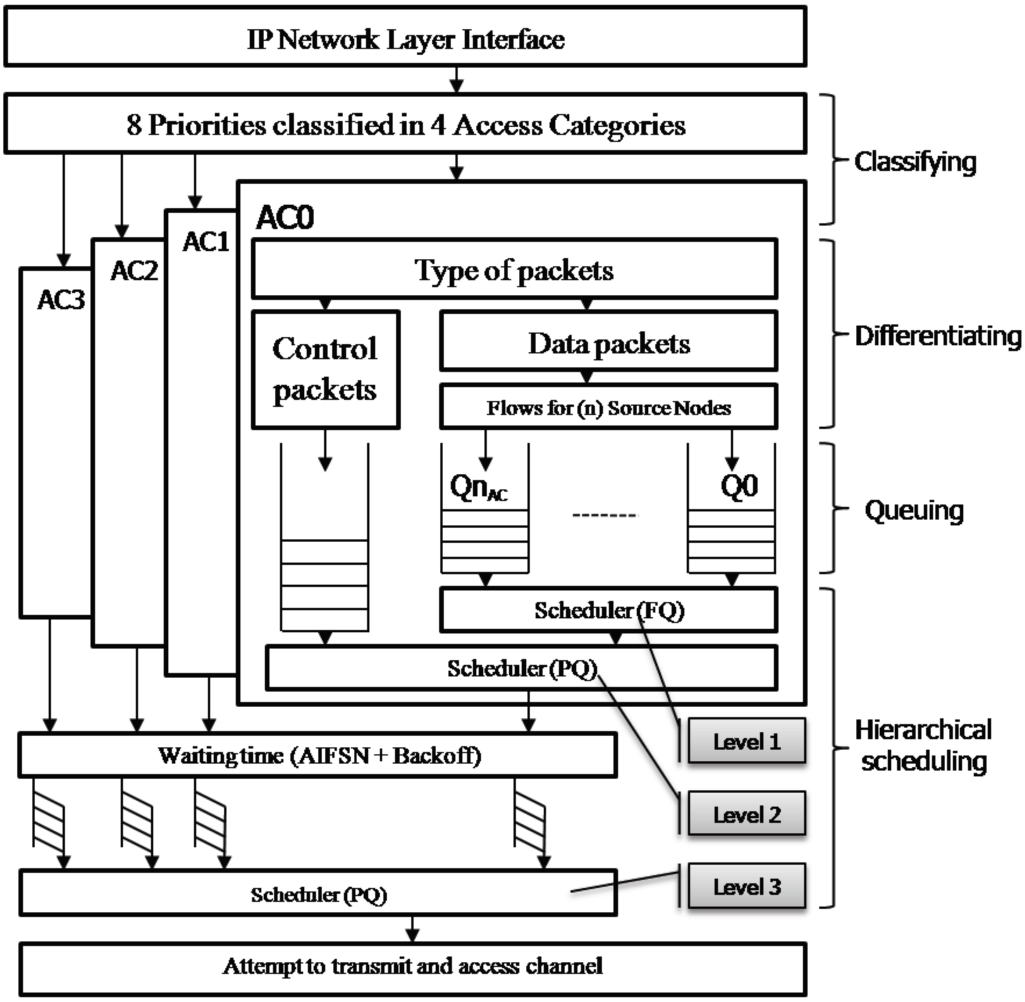


Figure 4.2: Model of FQ-EDCA

As seen in figure 4.2, each node station apply the following algorithm of FQ-EDCA in MAC layer:

- FQ-EDCA receives the packets from the upper layer (i.e. Network (IP) layer). Normally, these packets are prioritized according with the 8 priorities assigned in IP header. Then, it classifies the packets into 4 classes as defined in EDCA specifications (IEEE 05a).

- In each AC, it differentiates the packets, depending on their type, to control packets (RTS/CTS, Ack, IP routing, etc.) and data packets (TCP/UDP). This differentiation is important, to allow the control packets transmitted without delay. That leads to increase the efficiency of routing and QoS guarantee.

- Another differentiation is applied for the data packets, depending on their source nodes, to define one queue for each source node. However, in ad-hoc, the node may transmits its packets and the packets of its neighbors. That causes unfairness of resource allocation while congestion. Thereby, by this differentiation, the ill-behaved sources can be separated to be controlled.

- For queuing in each AC, one queue is defined for the control packets, while n_{ac_i} queues are defined for the data packets. However, n_{ac_i} is equal to the number of source nodes for AC_(ac) in this node (i). Let SN is equal to all the source nodes $\Rightarrow n_{ac_i} \subset SN$. If $i \in SN$ then $i \in n_{ac_i}$. If $i \notin SN$ then $i \notin n_{ac_i}$.

- A Fairness Queuing (FQ) mechanism (an extension of WFQ and WF2Q) is implemented. Where, every packet ($P^q_{k_q}$) of sequence (k_q) will be enqueued in queue (q) for ac_i. If queue (q) is full then this packet is dropped. However, queue (q) is defined for the source node (m) of this packet. So, $m \in n_{ac_i}, SN$. If $m \neq i$ then node(i) is routing the packets of node (m). FQ selects an eligible packet for ac_i. The eligible packet must pass through a hierarchical scheduling levels before attembting of transmission. This scheduling technique will be described in the next section.

4.2.2 Technical scheduling with fairness of FQ-EDCA

This scheduling technique is the basis of equity provided by the FQ-EDCA model. It ensures that all queues will be served in time, giving priority to those with the earliest packet in the lead (see the eligible packet notion).

In the FQ-EDCA model, as seen in figure 4.2, the following equations define the techniques of the first level of scheduling (Fairness Queuing (FQ)). Which is the most important part of FQ-EDCA's algorithm.

Let ac_i be an access category of the node i, where n_{ac_i} queues are effectives to schedule the packets in the node i. Each queue will be represented by the associated rank ($q \in [\![1, n_{ac_i}]\!]$). All the packets in a same queue are indexed by there rank. We will call $P^q_{k_q}$ the k^{th}_q packet in the queue q which is supposed to be at the head of the queue; all the previous packets ($\{P^q_1, \ldots, P^q_{k_q-1}\}$) have been sent. In this way, we define in the rest of this section a model which defines the techniques of the first level of scheduling.

The equations are described below; the expected start-number for serving the first byte of packet ($P^q_{k_q}$) queued at the head of queue (q). This value is given by the greatest value between the virtual-round-number (V) and the expected finish-number for the preceding packet ($F(P^q_{k_q-1})$).Thus we have:

$$\forall q \in [\![1, n_{ac_i}]\!], \forall k \in \mathbb{N}, S(P^q_{k_q}) = max(V, F(P^q_{k_q-1}))$$

Proof. The proof is based on the explanation previously given. This value must be greater than:

For all $q \in [\![1, n_{ac_i}]\!]$ and for all $k \in \mathbb{N}$

- The value of the virtual-round-number (V): $S(P^q_{k_q}) \geqslant V$
- The expected finish-number for the preceding packet ($F(P^q_{k_q-1})$): $S(P^q_{k_q}) \geqslant F(P^q_{k_q-1})$

Then we must have:

$$S(P^q_{k_q}) \geqslant max(V, F(P^q_{k_q-1}))$$

We want to optimize the system. In this context, we start as soon as possible. Naturally, the earliest date is given by the max of the previous value. Then, we obtain the next formula:

$$\forall q \in [\![1, n_{ac_i}]\!], \forall k \in \mathbb{N}, S(P^q_{k_q}) = \max(V, F(P^q_{k_q-1}))$$

$\square$

The expected finish-number is for serving the last byte of packet $(P^q_{k_q})$ queued at the head of queue (q) for ac_i. The expected finish-number will be given by the start number $(S(P^q_{k_q}))$ plus the size in byte of the packet $(L(P^q_{k_q}))$. Thus we have:

$$F(P^q_{k_q}) = S(P^q_{k_q}) + L(P^q_{k_q})$$

Proof. The expecting finish-number is, basically given by:

- The start-number: $S(P^q_{k_q})$
- The size of the packet $(P^q_{k_q})$: $L(P^q_{k_q})$

Thus we have:

$$F(P^q_{k_q}) = S(P^q_{k_q}) + L(P^q_{k_q})$$

$\square$

It will be useful to consider a special packet:

The eligible packet is the one that has the earliest finish-number, between packets that their start-number are below than the virtual-round-number. This packet will be abstract in the next by the equation below:

$$P_e = \underset{\{(k_q)|S(P^q_{k_q})\leqslant V\}}{\arg\min}\ {}^1F(P^q_{k_q})$$

[1]The function arg min stands for the argument of the minimum, that is to say, the set of points of the given argument for which the value of the given expression attains its minimum value. $\underset{x}{\arg\min} f(x) = \{x|\forall y, f(x) \leqslant f(y)\}$

Remark. P_e is a definition. The idea is to define a special packet which will be send in first. So we consider the set of all packet characterized by a start-number lower than the virtual-round-number. After that, we take the packet in the queue with the lower index.

To evaluate the virtual-round-number, we will need first the minimum value of the start-number between the n_{ac_i} queue for the access category and the V_{old} will be upgraded to a new value. This value could be either the minimum value of the start-number or increasing the V_{old} by a round number. This is described by the next equation:

$$S_{min} = \min_{1 \leqslant q \leqslant n_{ac_i}} \{S(P^q_{k_q})\}$$

$$V = max(V_{old} + \frac{L(P_e)}{n_{ac_i}}, S_{min})$$

Proof. The expecting virtual-round-number is, basically given by:

- $V \geqslant V_{old} + \frac{L(P_e)}{n_{ac_i}}$

  Normally, the round number is defined by the size of the served (eligible) packet. Thus, for each round, the next packet of this queue could be served after serving the other active queues $(n_{ac_i} - 1)$. That means, the packets queued at the head of n_{ac_i} queues could be served for each round.

- $V \geqslant S_{min}$ If the queue is empty after serving the packet.

Then we must have the equality below:

$$V \geqslant max(V_{old} + \frac{L(P_e)}{n_{ac_i}}, S_{min})$$

Then we set the equality for sending as soon as possible.

□

As a consequence, we obtain the next system of equations for our work. It describes by the following system:

$$S(P^q_{k_q}) = max(V, F(P^q_{k_q-1})) \tag{4.1}$$

$$F(P^q_{k_q}) = S(P^q_{k_q}) + L(P^q_{k_q}) \tag{4.2}$$

$$P_e = \underset{\{(k_q)|S(P^q_{k_q})\leqslant V\}}{\arg\min} F(P^q_{k_q}) \tag{4.3}$$

$$S_{min} = \min_{1\leqslant q\leqslant n_{ac_i}} \{S(P^q_{k_q})\} \tag{4.4}$$

$$V = max(V_{old} + \tfrac{L(P_e)}{n_{ac_i}}, S_{min}) \tag{4.5}$$

Where:

- $P^q_{k_q}$ is the first packet in the queue q. Whereas, all the previous packets $(\{P^q_1, \ldots, P^q_{k_q-1}\})$ have been sent.
- $S(P^q_{k_q})$ is the expected start-number for serving the first byte of packet $(P^q_{k_q})$ queued at the head of queue (q) for ac_i.
- $F(P^q_{k_q})$ is the expected finish-number for serving the last byte of packet $(P^q_{k_q})$ queued at the head of queue (q) for ac_i.
- $L(P^q_{k_q})$ is the size of that packet in bytes.
- V is the virtual-round-number (byte-by-byte serving mode) for ac_i. V_{old} is used to specify the previous value of the virtual round number.
- S_{min} is the minimum start-number between the (n_{ac_i}) queues for ac_i.
- n_{ac_i} is the number of queues for ac_i.
- P_e is the eligible packet, that has the earliest finish-number, between packets that their start-number are smaller than the virtual-round-number.

Moreover, FQ-EDCA's algorithm, that is described in the last section is continued by the following steps to select the eligible packet:

- At the initialization, $S(P_1^q)$ and V are equal to zero. However, $S(P_{k_q}^q)$ (4.1) and $F(P_{k_q}^q)$ (4.2) are calculated for each packet at the head of queue. Therefore V will be updated (4.5) after finding the minimum start-number between the non-empty queues of n_{ac_i} (4.4).
- The eligible packet must pass through a hierarchical scheduling levels.
  - For dequeuing, a Fairness Queuing (FQ) scheduler (first level) will serve the packet (P_e) queued at the head of queue that has the earliest finish-number, between packets that their start-number are below than the virtual-round-number (4.3).
  - Then in each AC, a Priority Queuing (PQ) scheduler (second level) will serve two types of outgoing packets (control and data packets). Thereby, depending on packet's type, the packet that has the highest priority is eligible. Normally, the control packet has the highest priority. Thus, the data packet will be served if the queue of the control packet is empty
  - The eligible packet at each AC waits its back-off and AIFS, depending on its AC. If several packets have the same wait time, a PQ scheduler (third level) will treat the elected packet depending on its priority. Thereby, the packet that has the highest priority is eligible and so on.

However, by this algorithm, the policy of IEEE802.11e (EDCA) is also respected. We will prove that by the simulations. Although, some of scheduling disciplines (i.e WFQ and WF2Q) are less complicated than FQ-EDCA, they need more calculations than the first level of scheduling of FQ-EDCA. Because in WFQ and WF2Q, there are weights distributed for each flow to be prioritzed. While in FQ-EDCA, these weights are implicitly defined by the classifications of priorities in each AC. That means, the parameters of EDCA (i.e. AIFSN, CWmin and CWmax) describe the weights of the flows in each AC.

Table 4.1: The serving sequence for each queued packet in the small example

Time	0	5	20	25	30	35
Head of queues	P11 P12	P21 P12	P21 P22	P31 P22	P41 P22	P51 P22
$Q_1(S,F)$	0,5	5,10	5,10	10,15	15,20	20,25
$Q_2(S,F)$	0,15	0,15	15,30	15,30	15,30	15,30
S_{min}	0	0	5	10	15	15
V	0	2.5	10	12.5	15	17.5
$S \leqslant V$	P11,P12	P12	P21	P31	P41,P22	P22
Serving pkt $\Rightarrow min(F)$	P11	P12	P21	P31	P41	P22

4.2.3 An illustration for FQ behavior

This section gives an illustrative example that allows understanding how FQ works. As seen in figure 4.3, let 2 queues are queuing packets of different sizes in whatever AC. At the input (queuing), the first queue is queuing 6 packets of size 5 units and the other is queuing 2 packets of size 15 units. Then, at the output of FQ (dequeuing), the packets will be dequeued and performed with a rate 1 unit/sec. So, the calculations of the output packets are performed by the equations (4.1,4.2,4.3,4.4,4.5). Therefore, the steps of the FQ's algorithm are arranged by the rows of table 4.1 for each serving time.

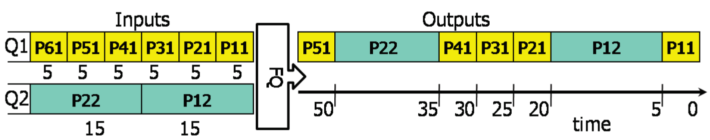


Figure 4.3: A small example explains the first level scheduling in FQ-EDCA

Now we will present the FQ-EDCA evaluations in multiple topologies and its performance, as followed by the sections below.

4.3 The simulations of FQ-EDCA in simple topologies

Network Simulator 2 (version 2.31 of ns-allinone (NS2)) is chosen to implement and evaluate FQ-EDCA. Therefore three scenarios are applied to evaluate FQ-EDCA compared with EDCA. Moreover, each scenario has a topology described in figure 4.4. These topologies are basic structures of ad-hoc networks. However, a cascaded nodes connected together to transmit their packets and forward the packets of their neighbors, as chain topology (i.e. topology (I)). And also, sometimes many of source nodes needs to forward their packets by the same routing node, as tree topology (i.e. topologies (II and III)). The las two topologies are proposed, as it will be expected, to present the difference of resource allocation btween the in-ranged source nodes and the out-ranged source nodes. Further, with increasing the out-ranged source nodes, the total bandwidth will be reduced because of increasing collisions and congestion. Thereby, these different behaviors cause different resource allocations between source nodes for each topology. They lead to unfairness as explained in the results of simulations (next section).

However, for each scenarios, nodes A, B, C and E (node E for topology III) are source nodes, while D is a sink node. D is implied to receive flows of A, B, C and E. Our aim in these simulations is to compare the respective bandwidth of the source nodes A, B, C and E. Therefore, each source node transmits 4 flows to D (one flow for each AC, i.e. AC_VO, AC_VI, AC_BE and AC_BK). The flows are sent at constant baud rate (constant bit rate) CBR/UDP traffic applications with constant packet size (see table 4.9). Further, in topologies (II and III), the distance between D and the center node is 250m to insure that only the centre node is in-ranged source node and the others are out-ranged source nodes. If that distance is made 200m, all the source nodes will be in-ranged source nodes else node E in topology (III). So, that thing couldnot give us the expected results.

In addition, FQ-EDCA is compared with EDCA model default parame-

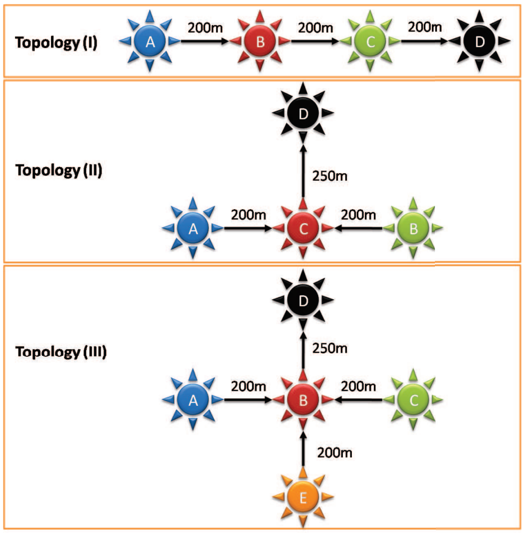


Figure 4.4: Topology (I) is a chain topology, where the source nodes(A,B and C) send data to D. Topology (II) is a T topology, where the source nodes(A,B and C) send data to D. And topology (III) is a tree topology, where the source nodes(A, B, C and E) send data to D.

ters (Wiet). To realize the comparison, bottlenecks and heavy loads of traffic flows should be applied in the network. For that reason, the CBR/UDP traffic is implied in each simulation. This is the case considered in the performed scenarios to show you the ability of FQ-EDCA for treating the congestion. All the parameters of simulations are presented in table 4.9.

4.4 The results of simulations

This section gives simulations results for the two models: FQ-EDCA and EDCA. Comparative simulations of the two models will enable us to evaluate the real benefits of FQ-EDCA. These simulations will be performed under topologies presented in the last section.

The results of our simulations are considered as following:

Table 4.2: The parameters of simulations

Standards	IEEE802.11b/e
QoS	EDCA
Topology mode	adhoc
Routing protocol	DSDV
Source nodes	A,B, E and C
Sink nodes	D
Number of flows	4
Number of ACs	4 (one flow for each AC)
Type of flows	CBR/UDP
Distance between nodes	200m
Transmission range	250m
Carrier sense range	550m
Packet size	1472 bytes
Packet baud rate	1Mbps
Bandwidth	11Mbps
Time of simulation	200s

4.4.1 The first scenario

As described, the first scenario is applied on a chain topology, as seen in figure 4.4 topology (I). The results of this scenario are given by figure 4.5. It describes the resource allocations for default EDCA model (Wiet) for different priorities. We remark that the allocated resources of bandwidth for A, B and C are different and lack the fairness for all the priorities. Indeed, C is queuing its packets and the routed packets of A and B in the same queue for each AC. Therefore, no control management is applied to control the ill-behavior of source nodes. So, the out-ranged sources (i.e. A and B) suffer from the unfairness of resource allocations.

Figure 4.6 describes the bandwidth allocations by using FQ-EDCA. We remark that the allocated bandwidth for the all sources are equal with respect to the priorities. Indeed, because of the robustness and the strength of FQ-EDCA, it can separate the ill-behaved sources. As well, we observe that the allocated bandwidth for all priorities are equal. That leads to minimize the delay of the out-ranged sources with balance, as seen in table 4.3. The delay is calculated from the sender to the receiver. Moreover, FQ-EDCA balances the average lost packet rate, as seen in table 4.4. The values of average lost

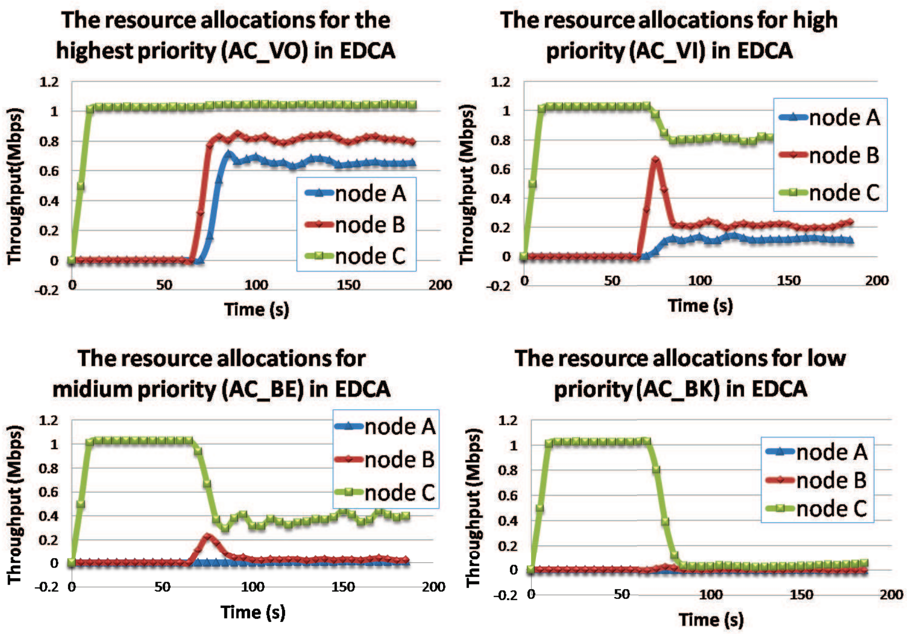


Figure 4.5: EDCA default behavior in the first scenario

Table 4.3: Delay comparison between EDCA vs. FQ-EDCA in Scenario I

	EDCA			FQ-EDCA		
AC	A	B	C	A	B	C
AC_BK	-	1.8s	0.19s	1.2s	1.1	1.1
AC_BE	2.6s	0.33s	20ms	73ms	70ms	67ms
AC_VI	55ms	32ms	8ms	20ms	20ms	20ms
AC_VO	9ms	8ms	6ms	7ms	7ms	7ms

packet rate, that are presented in the table, are measured for each sample time of 5sec withen the simulation.

Further, for the lowest priority (AC_BK), it is clear that there is no resources available for this AC in both of FQ-EDCA and EDCA, because all the bandwidth is allocated for the higher priorities. Therefore, to show you the state of starvation and bottleneck occurred, we are compelled to present the behavior of lowest priority, as seen in figure 4.5 and figure 4.6.

As a result, in the first scenario, FQ-EDCA enhances EDCA and guarantees services by implementing the fairness, see figure 4.7. The total bandwidth of the network is nearly 4.2Mbps for both EDCA and FQ-EDCA.

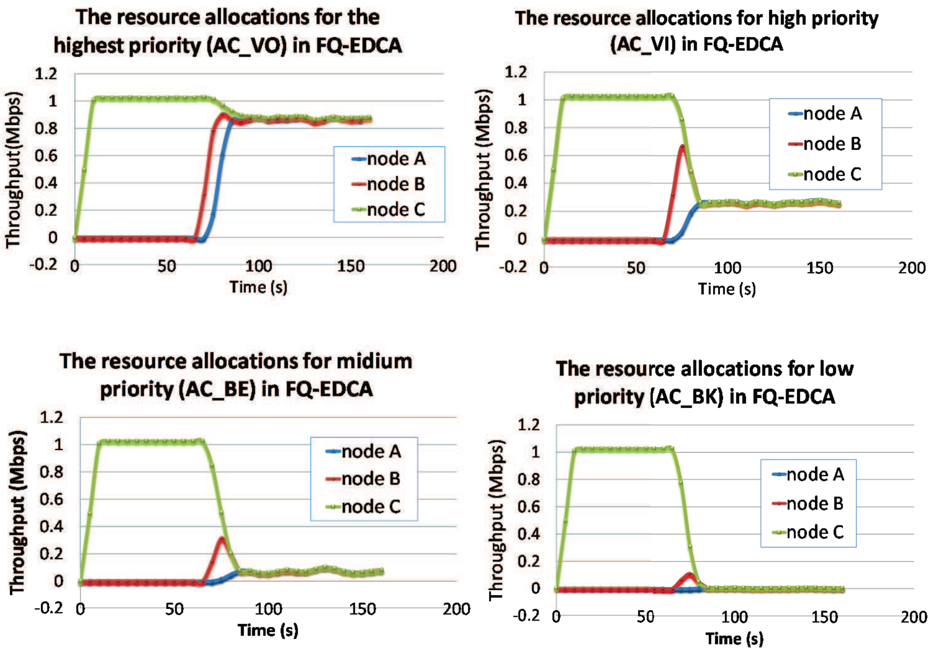


Figure 4.6: FQ-EDCA behavior in the first scenario

Table 4.4: The average lost packet rate between EDCA vs. FQ-EDCA in Scenario I

	EDCA			FQ-EDCA		
AC	A	B	C	A	B	C
AC_BK	100%	100%	97%	98%	98%	98%
AC_BE	100%	97%	80%	96%	96%	96%
AC_VI	92%	89%	40%	84%	84%	84%
AC_VO	53%	40%	0%	26%	26%	25%

That means FQ-EDCA has almost the same complexity as EDCA with no affection on the total bandwith or network throughput. Further, FQ-EDCA enhances EDCA by allocating the resources with fairnes for all source nodes of the same priority class.

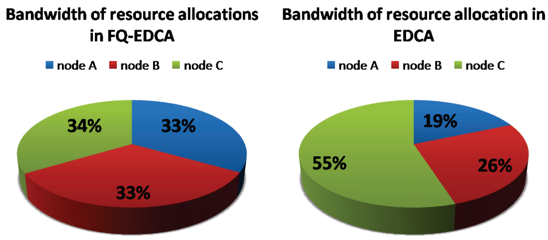


Figure 4.7: Total Bandwidth of resource allocations in scenario (I) for both FQ-EDCA and EDCA

4.4.2 The second scenario

As described, the second scenario is applied on a star or T topology, as seen in figure 4.4 topology (II). The results of this scenario are given by figure 4.8. It explains the described phenomenon previously in the first scenario. We remark that the allocated resources of bandwidth for A and B are less than C for all the priorities. As described, EDCA lacks the fairness and has no control management on the ill-behaved sources. So, the out-ranged sources (i.e. A and B) suffer from the unfairness of resource allocations.

Figure 4.9 describes the bandwidth allocations by using FQ-EDCA. We remark that the allocated bandwidth for all sources are equal with respect to the priorities. As described in the first scenario, the robustness of FQ-EDCA can separate the ill-behaved sources. That leads to decrease with balance the delay of the out-ranged sources, as seen in table 4.5. And also, the average lost packet rate is balanced, as seen in table 4.6. The values of average lost packet rate, that are presented in the table, are measured for each sample time of 5sec withen the simulation.

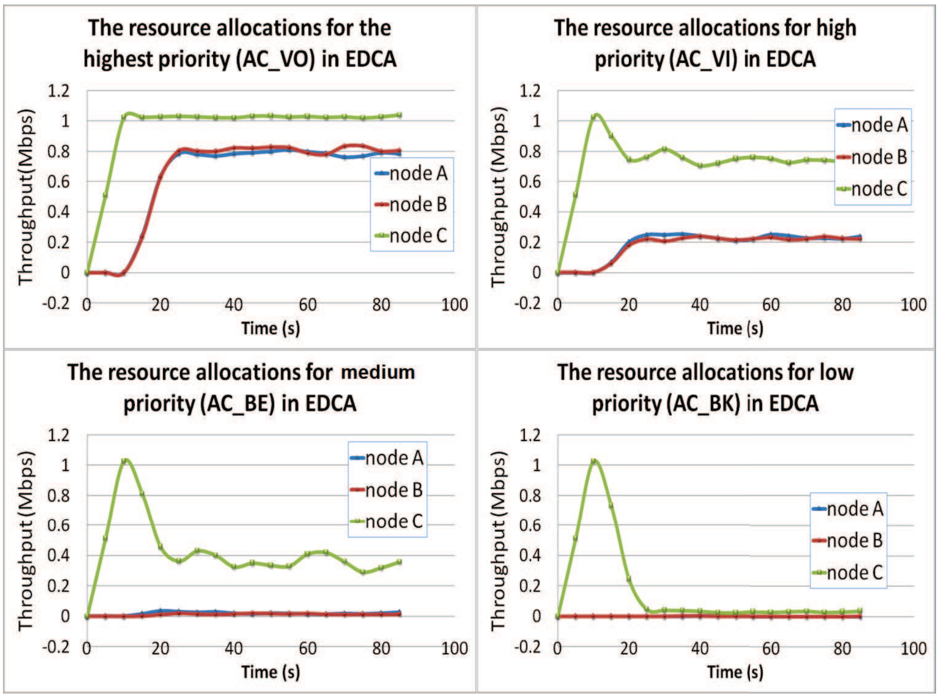


Figure 4.8: EDCA default behavior in the second scenario

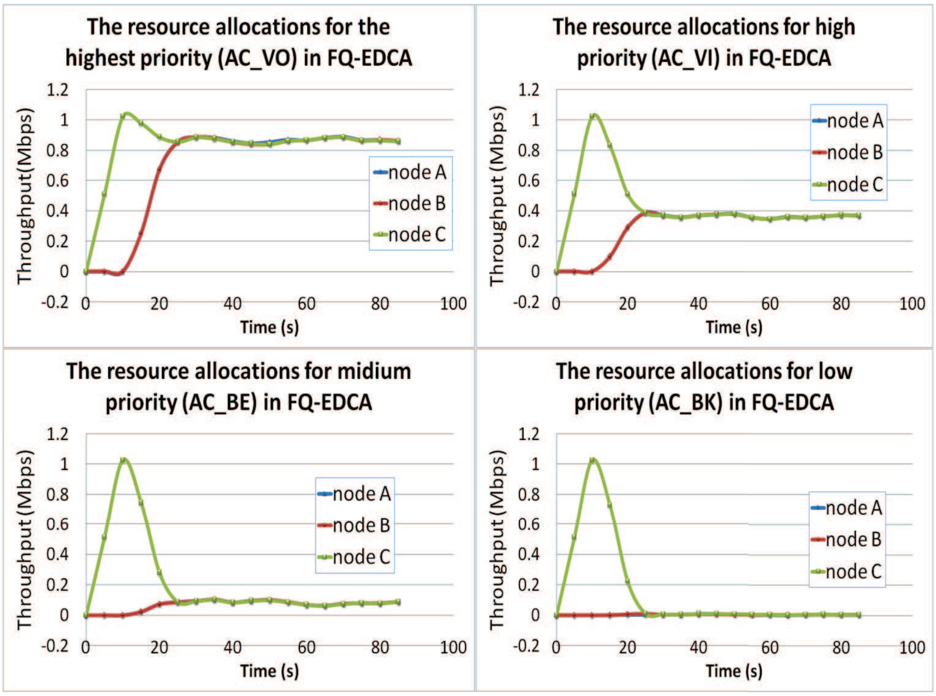


Figure 4.9: FQ-EDCA behavior in the second scenario

Table 4.5: Delay comparison between EDCA vs. FQ-EDCA in Scenario II

	EDCA			FQ-EDCA		
AC	A	B	C	A	B	C
AC_BK	-	-	0.2s	1.1s	1.1s	1s
AC_BE	0.42s	0.48s	20ms	78ms	78ms	73ms
AC_VI	31ms	31ms	9ms	17ms	17ms	16ms
AC_VO	8ms	8ms	6ms	7ms	7ms	7ms

Table 4.6: The average lost packet rate between EDCA vs. FQ-EDCA in Scenario II

	EDCA			FQ-EDCA		
AC	A	B	C	A	B	C
AC_BK	100%	100%	97%	98%	98%	98%
AC_BE	99%	99%	79%	94%	94%	94%
AC_VI	89%	89%	45%	78%	78%	78%
AC_VO	41%	40%	0%	26%	26%	26%

Further, for the lowest priority (AC_BK), it is clear that there is no resources available for this AC in both of FQ-EDCA and EDCA, because all the bandwidth is allocated for the higher priorities. Therefore, to show you the state of starvation and bottleneck occurred, we are compelled to present the behavior of lowest priority, as seen in figure 4.8 and figure 4.9.

As a result, in the second scenario, FQ-EDCA enhances EDCA and guarantees services by implementing the fairness, see figure 4.10. The total bandwidth of the network is nearly 4.1Mbps for both EDCA and FQ-EDCA. That means FQ-EDCA has almost the same complexity as EDCA with no affection on the total bandwith or network throughput. Further, FQ-EDCA enhances EDCA by allocating the resources with fairnes for all source nodes of the same priority class.

4.4.3 The third scenario

As described, the third scenario is applied on a tree topology, as seen in figure 4.4 topology (III). The results of this scenario are given by figure 4.11. It explains the described phenomenon previously in the second scenario. We remark that the allocated resources of bandwidth for A, C and E are less

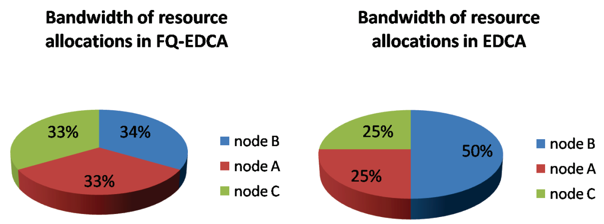


Figure 4.10: Total Bandwidth of resource allocations in scenario (II) for both FQ-EDCA and EDCA

Table 4.7: Delay comparison between EDCA vs. FQ-EDCA in Scenario III

	EDCA				FQ-EDCA			
AC	E	A	B	C	E	A	B	C
AC_BK	-	-	0.7s	-	3s	3s	3s	3s
AC_BE	1.5s	1.5s	0.4s	1.5s	0.2s	0.2s	0.2s	0.2s
AC_VI	25ms	25ms	6ms	25ms	18ms	18ms	18ms	18ms
AC_VO	18ms	18ms	6ms	18ms	13ms	13ms	13ms	13ms

than B for all the priorities. However, EDCA still lacks the fairness and has no control on the ill-behaved sources. So, the out-ranged sources (i.e. A, C and E) suffer from the unfairness of resource allocations.

Figure 4.12 describes the bandwidth allocations by using FQ-EDCA. We remark, as expected, that the allocated bandwidth for the all sources are equal with respect to the all priorities. As described in the second scenario, the robustness of FQ-EDCA can separate the ill-behaved sources. This leads to decrease with balance the delay of the out-ranged sources, as seen in table 4.7. And also, the average lost packet rate is balanced, as seen in table 4.8. The values of average lost packet rate, that are presented in the table, are measured for each sample time of 5sec withen the simulation.

Further, for the lowest priority (AC_BK), it is clear that almost there is no resources available for this AC in both of FQ-EDCA and EDCA, because all the bandwidth is allocated for the higher priorities. Therefore, to show you the state of starvation and bottleneck occurred, we are compelled to present the behavior of lowest priority, as seen in figure 4.11 and figure 4.12.

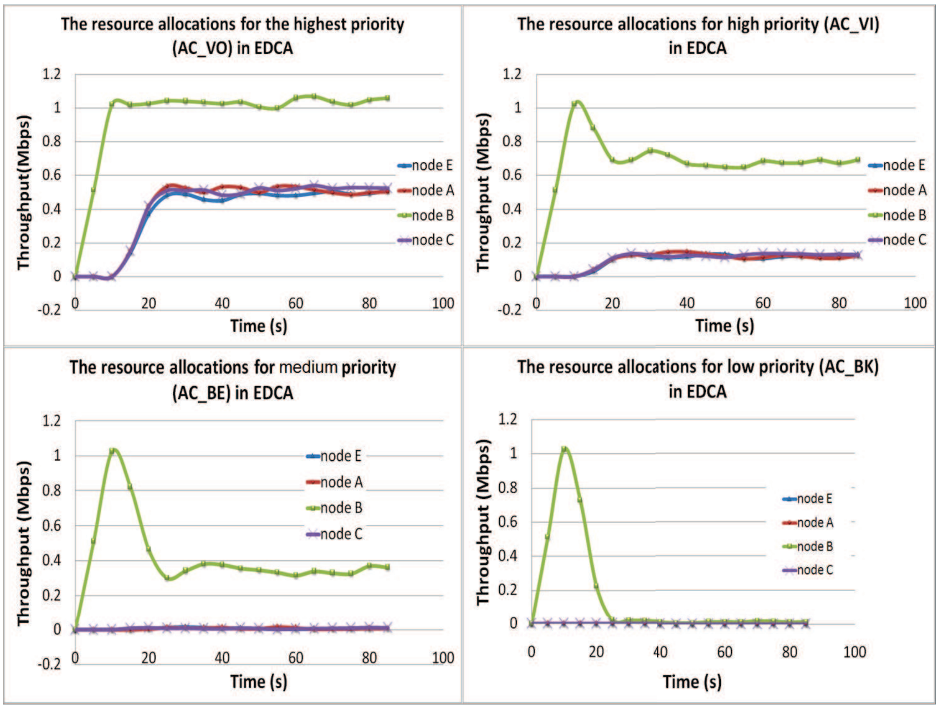


Figure 4.11: EDCA default behavior in the third scenario

Table 4.8: The average lost packet rate between EDCA vs. FQ-EDCA in Scenario III

	EDCA				FQ-EDCA			
AC	E	A	B	C	E	A	B	C
AC_BK	100%	100%	100%	100%	100%	100%	100%	100%
AC_BE	100%	100%	81%	100%	98%	98%	98%	98%
AC_VI	93%	92%	50%	92%	85%	85%	85%	85%
AC_VO	74%	74%	0	74%	63%	63%	63%	63%

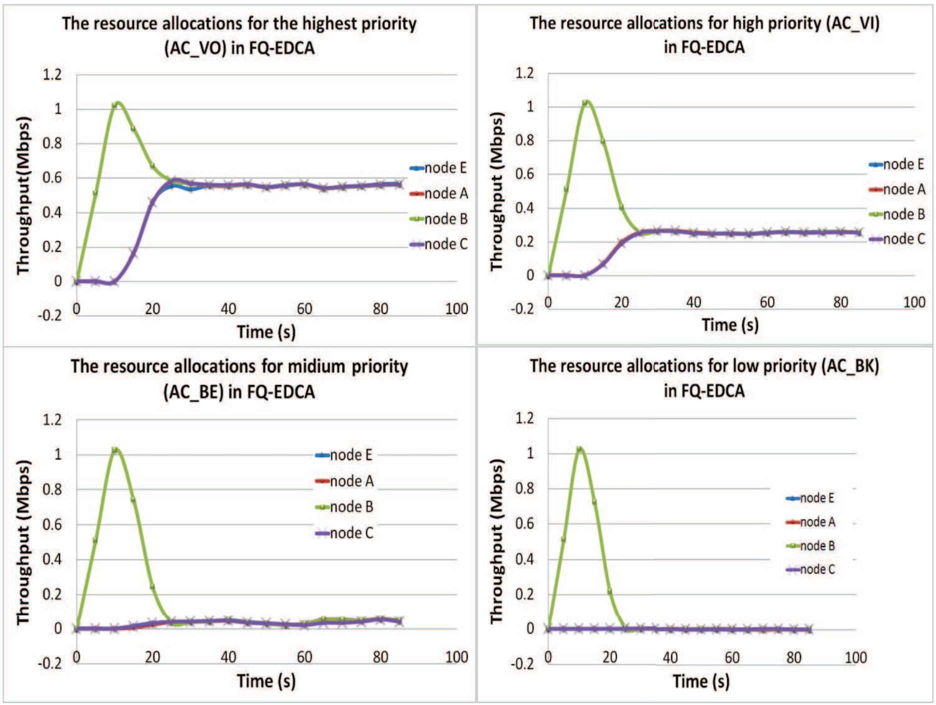


Figure 4.12: FQ-EDCA behavior in the third scenario

As a result, in the third scenario, FQ-EDCA enhances EDCA and guarantees services by implementing the fairness, see figure 4.13. The total bandwidth of the network is nearly 3.5Mbps. That means FQ-EDCA has almost the same complexity as EDCA with no affection on the total bandwith or network throughput. Further, FQ-EDCA enhances EDCA by allocating the resources with fairnes for all source nodes of the same priority class. And also, as expected for the difference between topology II and III, the total bandwidth of the network is reduced because of collisions and congestion .

As a result, in these scenarios, FQ-EDCA enhances EDCA and guarantees services by implementing the fairness. Thereby, the QoS is optimized with FQ-EDCA in simple topologies.

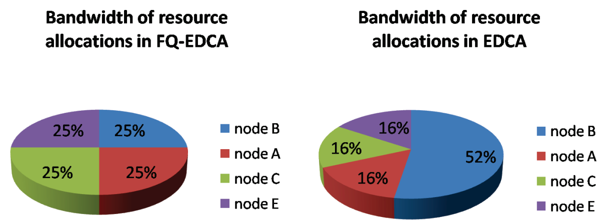


Figure 4.13: Total Bandwidth of resource allocations in scenario (III) for both FQ-EDCA and EDCA

4.5 An extension of FQ-EDCA simulations

This extension aims to evaluate FQ-EDCA and its strength in large topologies (i.e. 50 to 100 nodes) that could be applied in real ad-hoc WLANs. FQ-EDCA enhances the QoS of EDCA by implementing a hierarchical scheduling techniques and improving the architecture of EDCA scheme. Thereby, the traffic is regulated between the source nodes. Thus, the bandwidth is distributed fairly between all the source nodes. As a consequence, these nodes can transmit their packets fairly with respect to the EDCA specifications. Moreover, the end-to-end delay of packets for the flows of the same priorities is bounded (i.e. for out-ranged source nodes). As expected, the services guarantees of QoS is enhanced by FQ-EDCA, which is proved and analyzed by experimental simulations for large topologies.

However, many of simulations are applied to evaluate FQ-EDCA compared with EDCA. Indeed, This comparison could be sufficient between these models, because it focuses the benefits earned by our model over EDCA. And as known, EDCA is the base and the most popular standard of QoS in wireless networks. Moreover, these simulations are performed in a large topology described in figure 4.14. This topology is considered as a general structure of ad-hoc network.

However, the nodes are distributed in organized form, that are described as the following levels of hops:

- At level (0): 1 sink node (destination) in the center.
- At level (1): 3 source nodes in the range of destination (In-ranged nodes).
- At level (2): 6 source nodes out of range of destination (Out-ranged nodes).
- At level (3): 12 source nodes out of range of destination (Out-ranged nodes).
- At level (4): 24 source nodes out of range of destination (Out-ranged nodes).

More than these nodes, the network will be died with taking into account the simulations parameters described in table 4.9. However, every level has a distance nearly of 200m faraway of the level below, and each node has a range distance equal to 250m. Normally, we choose this values because by default the simulation (NS-2) use this range distance depending on the physical parameters (i.e. signal power, line of site, interfernce, etc.).

Our aim in these simulations is to compare the respective bandwidth for each source node. Therefore, each source node transmits to the sink node, as shown in figure 4.14, 3 flows at the same time (one flow for each AC: AC_VO, AC_VI and AC_BE). The flows are sent at constant baud rate (constant bit rate) CBR/UDP traffic applications with constant packet size (see table 4.9). Intentionally the CBR traffics are performed for multimedia and Best-effort data to obtain the worst-case of congestion and make our results be evaluated.

In addition, FQ-EDCA is compared with EDCA model default parameters (Wiet). To realize the comparison, many simulations are performed depending on the level of source nodes around the sink node (i.e. one level $\Rightarrow$ 3, two levels $\Rightarrow$ 9, three levels $\Rightarrow$ 21 and four levels $\Rightarrow$ 45 source nodes). This is the case considered in the performed simulations to show you the

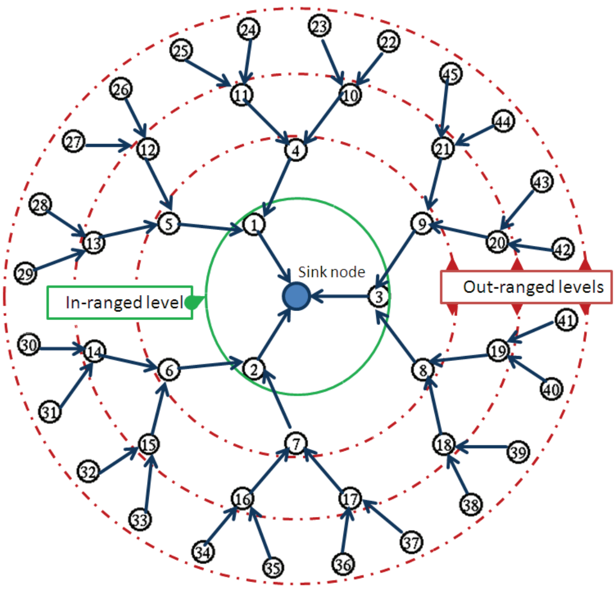


Figure 4.14: Topology of (45) source nodes organized with multi-hop distribution around (1) sink node.

ability of FQ-EDCA for treating the congestion in large topologies. All the parameters of simulations are presented in table 4.9.

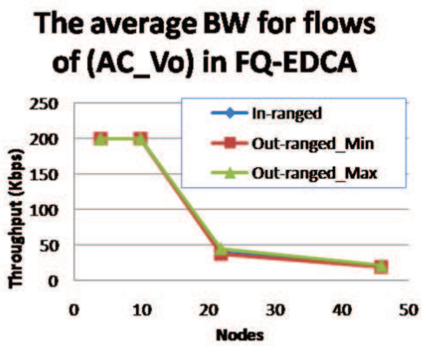
4.5.1 Results of simulations

The results of the last simulations obviously illustrate the problem of unfairness in EDCA. This unfairness is because, the nodes must cooperate to be connected together in multi-hop mode to transmit their packets and forward (route) the packets of their neighbors. So, the data packets received from upper layer are defined as owned packets. And those from PHY layer are defined as neighbors (routed) packets. Indeed in MAC layer, these packets are queued in one queue for each AC at the source node. However, because of the problems of wireless networks as explained, the owned packets are received to be queued faster than the routed packets. That gives more of the chance to access channel for the owned packets. As a result, under these conditions, the in-ranged nodes will consume the most of allowed bandwidth.

Table 4.9: The parameters of simulations

Standards	IEEE802.11b/e
QoS	EDCA
Topology mode	ad-hoc
Routing protocol	DSDV
Source nodes	$3, 9, 21$ and 45
Sink nodes	1
Type of flows	CBR/UDP
Distance between nodes	$\approx 200m$
Transmission range	$250m$
Carrier sense range	$550m$
Packet size	$1472bytes$
Packet baud rate	$200kbps$
Bandwidth	$11Mbps$

That means, as illustrated in figure 4.15, the allocated bandwidth of voice traffic (i.e. AC_Vo priority) for the in-ranged source nodes is more than the allocated bandwidth of the same traffic for the out-ranged source nodes.



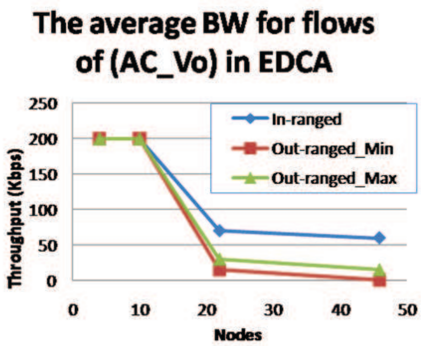


Figure 4.15: Comparison between FQ-EDCA and EDCA for Voice traffic priority

And also, in figure 4.16, the allocated bandwidth of video traffic (i.e. AC_Vi priority) for the in-ranged source nodes is more than the allocated bandwidth of the same traffic for the out-ranged source nodes.

Further, in figure 4.17, the allocated bandwidth of best-effort traffic (i.e. AC_Be priority) for the in-ranged source nodes is more than the allocated bandwidth of the same traffic for the out-ranged source nodes.

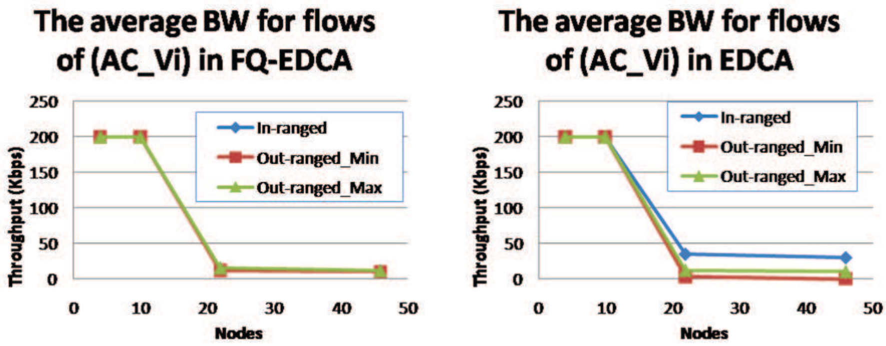


Figure 4.16: Comparison between FQ-EDCA and EDCA for Video traffic priority

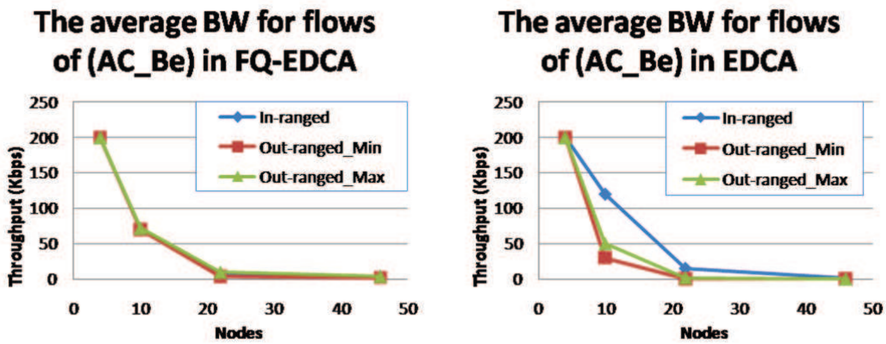


Figure 4.17: Comparison between FQ-EDCA and EDCA for Best-effort traffic priority

Table 4.10: Delay comparison between EDCA vs. FQ-EDCA

		EDCA			FQ-EDCA		
AC	Nodes	In-ranged	Out-ranged Min	Out-ranged Max	In-ranged	Out-ranged Min	Out-ranged Max
VO	3	30ms	30ms	30ms	30ms	30ms	30ms
	9	30ms	30ms	30ms	30ms	30ms	30ms
	21	50ms	0.2s	0.5s	50ms	50ms	0.2s
	45	70ms	0.1s	7s	0.3s	0.3s	0.6s
VI	3	30ms	30ms	30ms	30ms	30ms	30ms
	9	30ms	30ms	30ms	30ms	30ms	30ms
	21	0.2s	0.5ss	2s	0.5s	0.5s	0.7s
	45	0.6s	1s	25s	2s	2s	2.5s
BE	3	30ms	30ms	30ms	30ms	30ms	30ms
	9	50ms	0.1s	0.2s	80ms	90ms	0.1s
	21	0.3s	1s	30s	1s	1s	1.6s
	45	-	-	-	2s	2s	5s

Whereas, the difference of bandwidth between the out-ranged source nodes is bounded by the curves (Out-ranged_Min, Out-ranged_Max) for each traffic priority.

Logically, for the same problem explained above, the allocated bandwidth for the out-ranged source nodes of level (2) is more than that of level (3). And also, the allocated bandwidth for the source nodes of level (3) is more than that of level (4), and so on for each traffic priority.

Therefore, these different resource allocations between source nodes lead to the unfairness explained, which is resolved by FQ-EDCA.

Indeed, because of the strength of FQ-EDCA, it can separate the ill-behaved sources. As well, we observe that the allocated bandwidth for all priorities are equal. That leads to minimize the end-to-end delay of the out-ranged sources with balance, as seen in table 4.10. And also, FQ-EDCA balances the average lost packet.

As a result, in these simulations, FQ-EDCA enhances EDCA and guarantees services by implementing the fairness in large topologies. Thereby, the QoS is optimized with FQ-EDCA in these topologies. As well, this will be shown in the next section.

4.6 Network Performance of FQ-EDCA

FQ-EDCA model has been developed and implemented using network simulator (ns-2) to study the throughput, average delay and fairness performance, and to compare them with the recent standards of wireless networks (i.e. the IEEE802.11e (EDCA)).

Before evaluating the network performance of our model, we will define the performance metrics.

4.6.1 Performance Metrics

Several metrics can be defined to grade the performance of our model (FQ-EDCA) against the recent standards (IEEE802.11e) in ad-hoc wireless networking for QoS (EDCA). Therefore, some of these metrics have been carefully chosen to give an idea of the behavior and the reliability of FQ-EDCA in ad-hoc wireless networks.

However, these metrics measure the data transmission's features with respect to data packets. As a consequence, the analysis focuses on the benefits obtained by FQ-EDCA compared with other models. An explanation of these metrics follows:

Network Throughput

It is a measure of the amount of data (Bytes) transmitted from the source to the destination in a unit period of time (seconds). Considering the data rates and throughputs supported by IEEE802.11 standard, the throughput is measured in total bits received per second. Also, this metric only measures the total data throughput over the network.

The throughput of a node is measured by first counting the total number of data packets successfully received at the node, and computing the number of bits received, which is finally divided by the total simulation runtime.

The throughput of the network is finally defined as the average of the throughput of all nodes. Therefore, throughput can be stated as (4.6):

$$Throughput\ of\ a\ Node = \frac{Total\ Received\ Data\ Bits}{Simulation\ Runtime} \tag{4.6}$$

Similarly, the throughput for the network can be described as (4.7):

$$Network\ Throughput = \frac{Sum\ of\ Nodes'\ Throughput}{Number\ of\ Nodes} \tag{4.7}$$

End-to-End Delay

The end-to-end delay is the time taken for a data packet to reach the destination node. The term "End-to-End" means from an end-point sender to an end-point receiver.

When the simulations are performed with a random topology, the destination, that has the maximum number of hops from the sender, receives data packet with maximum delay. Therefore, the delay can be stated as (4.8):

$$PacketDelay = \underbrace{Received\ Time}_{at\ Destination} - \underbrace{Transmitted\ Time}_{at\ Source} \tag{4.8}$$

Moreover, the delay for a packet is the time taken for it to reach the destination. And the average delay is calculated by taking the average of delays for every data packet transmitted successfully. Thus, the average delay of the network can be stated as (4.9):

$$AverageDelay = \frac{Sum\ of\ all\ Packet\ Delays}{Total\ Number\ of\ Received\ Packets} \tag{4.9}$$

Delivery Ratio

It refers to the percentage of the transmitted data packets that are successfully received over all sent packets. The all sent packets are considered as; the summation of the dropped packets and the received packets.

However, it is an important metric, which can be used as an indicator to a congested network. In other words, the delivery ratio is calculated as the percent of received packets to the sent packets. Therefore, the delivery ratio can be calculated as (4.10):

$$Delivery\ Ratio = \frac{Number\ of\ Received\ Packets}{Number\ of\ Transmitted\ Packets} * 100\% \tag{4.10}$$

And the same thing for the delivery ratio of the network as stated in (4.11):

$$NetworkDeliveryratio = \frac{Sum\ of\ Delivery\ ratios}{Number\ of\ Nodes} \tag{4.11}$$

Another metric of the network performance is the energy consumption. However, in our studies is neglected, because we focuses our studies in the network performance for data transmission.

Now, we will analyze the performance of FQ-EDCA.

4.6.2 The performance analysis of FQ-EDCA

To evaluate the performance of FQ-EDCA, we will implement it in real topologies with different number of nodes and different data transmission rates. Thus, random topologies are applied for this evaluation. These topologies are generated by using topology generation tool (built on the random waypoint algorithm) implemented in the network simulator (NS2).

However, as seen in figures below, multiple topologies are applied for different number of nodes distributed randomly to achieve real benefits and make FQ-EDCA applicable in the real worlds. Also, these topologies are applied for two purpose. The first is for the network performance measurements, and the second is for the end-to-end performance measurements.

4.6.3 The performed simulations

The simulations are performed for last topologies (from 5 to 50 nodes). However, these nodes are enumerated from (0) to $(N-1)$, where N is the number of nodes for each topology. Thereby, the number of source nodes equals to

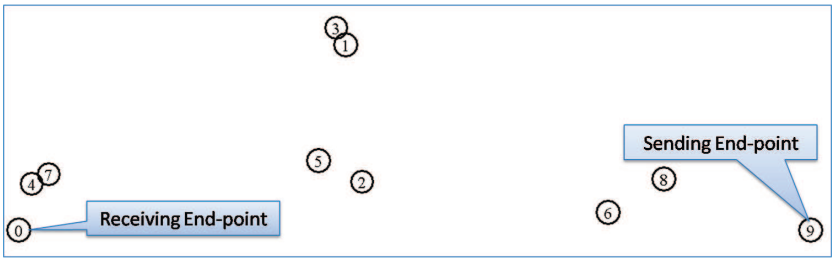


Figure 4.18: A random topology of 10 nodes

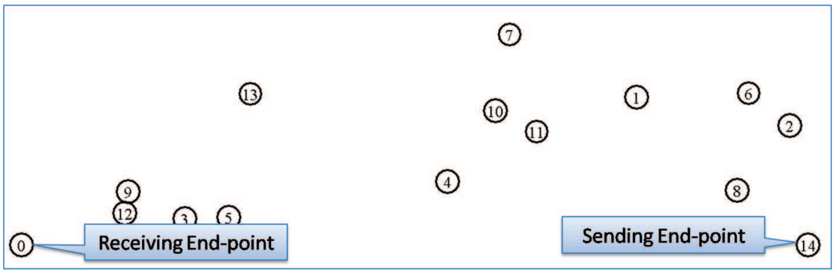


Figure 4.19: A random topology of 15 nodes

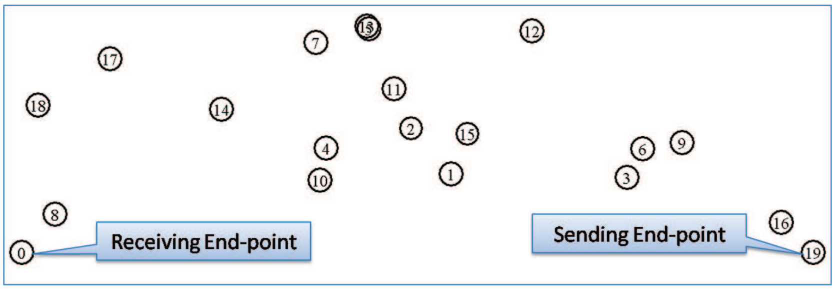


Figure 4.20: A random topology of 20 nodes

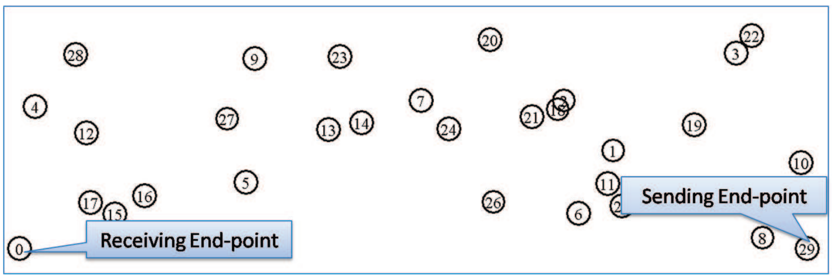


Figure 4.21: A random topology of 30 nodes

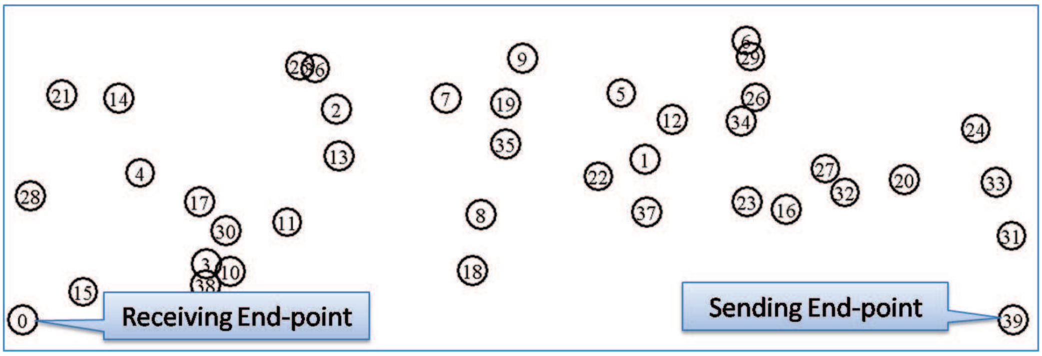


Figure 4.22: A random topology of 40 nodes

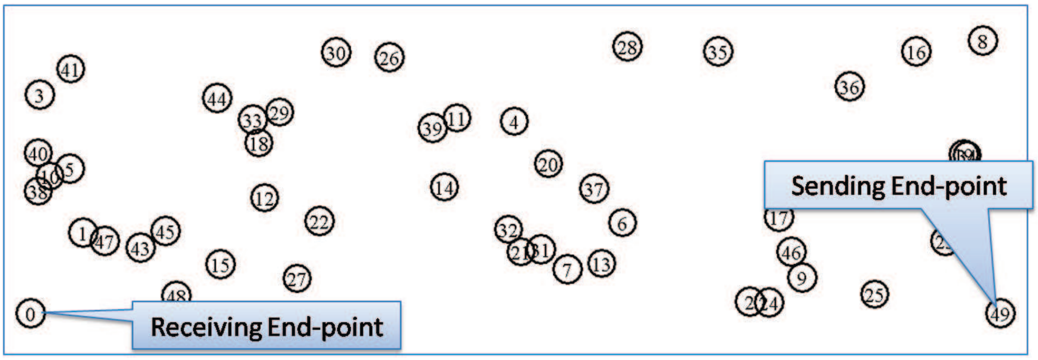


Figure 4.23: A random topology of 50 nodes

Table 4.11: The parameters of performed simulations

Standards	IEEE802.11b/e
QoS	EDCA & FQ-EDCA
Topology mode	ad-hoc
Routing protocol	DSDV
Source nodes	$5 - 50$
Type of flows	CBR/UDP
Transmission range	$250m$
Carrier sense range	$550m$
Packet size	$512bytes$
Packet baud rate	$10 - 100kbps$
Bandwidth	$1Mbps$

$(N - 1)$ for each topology. Every source node sends its packets to a random destination with a constant baud rate. Whereas, the baud rate is the user sending rate in bits per seconds. However, the simulations are repeated for different baud rates, see table 4.11. So, high baud rates make the network behavior overloaded. That leads to a congestion in the network. Also, this leads to increase the ill-behaved sources which consume the most of network resources.

All the nodes are distributed randomly except 2 nodes, which are enumerated with 0 and $(N - 1)$. These nodes are considered as the end-points, marked as receiving end-point and sending end-point. This distribution is made intentionally to obtain the end-to-end measurements. These measurements are essential for end-to-end QoS comparison.

The parameters of performed simulations are stated in table 4.11. However, these simulations are applied with the same parameters for both EDCA and FQ-EDCA, to compare their results.

4.6.4 The results of the performed simulations

After the simulations, many of measurements are obtained to analyze the performance of FQ-EDCA and compare it with EDCA. the results are obtained as following:

Network Throughput

As described, the network throughput is a metric of the network performance, which can describe the general state of the network traffic.

Thus, as seen of the curves of EDCA in figure 4.24, we remark that, in EDCA, the network throughput increases when increasing the baud rate until the source nodes consume the most resources of the network. And also, it is clear that when the number of nodes is high the network throughput is minimized, because the number of collisions and retransmissions are increased. Thus, for large number of source nodes with heavy load, this degrades the network throughput.

On the other hand, as seen of the curves of FQ-EDCA in figure 4.24, we remark that, in FQ-EDCA, the Network throughput is increased with increasing the baud rate until the source nodes consume the most resources of the network. And also, it is clear that FQ-EDCA enhances the network throughput of EDCA specially for small number of source nodes. Also, this enhancement is minimized for increasing the number of source nodes, because the number of collisions and retransmissions is still high.

Moreover, in network throughput analysis, we cannot conclude the benefits of FQ-EDCA over EDCA, because in both models they utilize the available resources of the network. Further, it does not take into account the differentiating between routed and owned packets and the fairness. While, the main enhancement of FQ-EDCA is to apply fairness. For that reason, they have nearly the same average network throughputs, specially for the network throughputs of 20 nodes and above.

In addition, in FQ-EDCA, almost all the source nodes have the chance to send their packets. While, $(30 - 50\%)$ of source nodes are dead nodes in EDCA, specially for large number of source nodes and high baud rate. Thus, the source nodes of one-hop destination (in-ranged nodes) have the most of resource allocations than the source nodes of multi-hop destination (out-ranged nodes) in EDCA. While, the fairness of resource allocation is applied between the source nodes of one-hop and multi-hop destination in

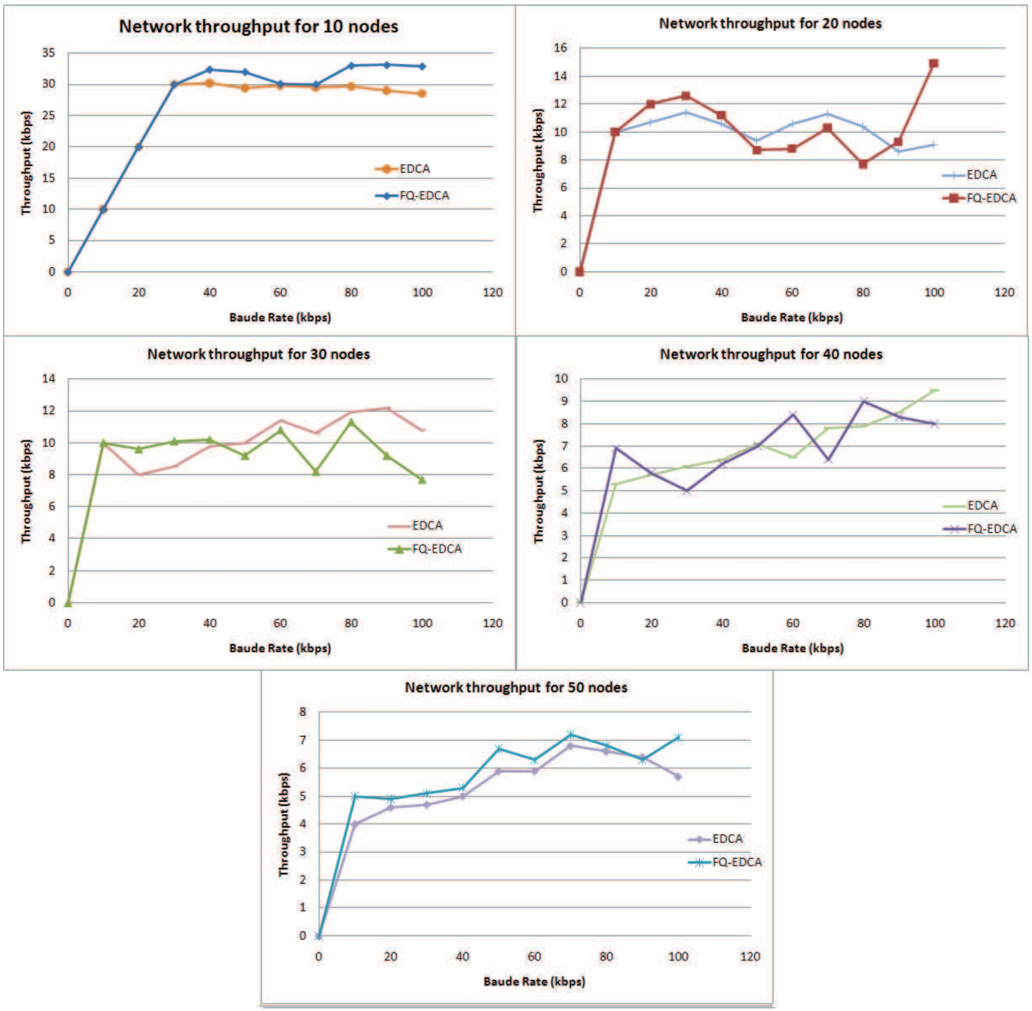


Figure 4.24: The network throughputs of EDCA and FQ-EDCA for 10, 20, 30, 40 and 50 nodes

FQ-EDCA

Indeed, in general for both FQ-EDCA and EDCA, they have the same phenomenon of network throughput. That means, the complexity of FQ-EDCA is still like EDCA with taking into account the enhancement that will be obtained for other metrics of performance.

Further, let us present the end-to-end network performance.

End-to-end Throughput

The end-to-end throughput is an important metric of the network performance and QoS, which can describe the end-to-end throughput for the longest path traffic. This path is established between sending and receiving end-points.

Thus, as seen of the curves of EDCA in figure 4.25, we remark that, in EDCA, the end-to-end throughput is increased with increasing the baud rate until the source nodes consume the most resources of the network. Then, the end-to-end throughput is decreased, because of the lack of resources. And also, it is clear that when the number of source nodes is increased, the end-to-end throughput is minimized, because the number of collisions and retransmissions are increased. Thus, for large number of source nodes with heavy load, the number of ill-behaved sources is increased. Therefore, this degrades the end-to-end throughput and makes it reach to zero.

While, as seen of the curves of FQ-EDCA in figure 4.25, we remark that the end-to-end throughput in FQ-EDCA is more stable and robust. This makes the QoS guarantee is increased. So, it is clear that FQ-EDCA enhances EDCA and the end-to-end throughput.

End-to-end delay

The end-to-end delay is an important metric of the network performance and QoS, which can describe the worst case path delay between sending and receiving end-points.

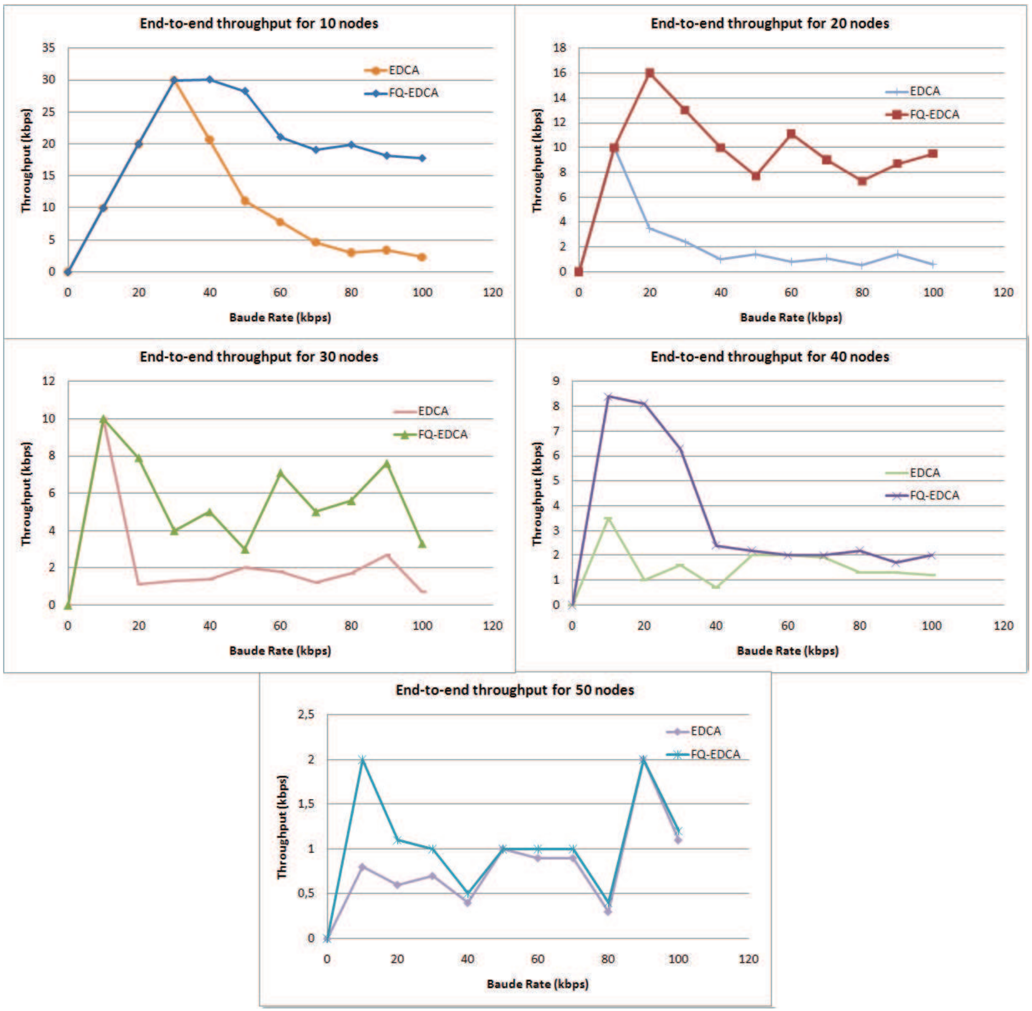


Figure 4.25: The end-to-end throughput in EDCA and FQ-EDCA for 10, 20, 30, 40 and 50 nodes

As seen of the curves of EDCA in figure 4.26, we remark that in EDCA, the end-to-end delay is increased exponentially with increasing the baud rate of senders. This delay becomes infinity (dead link) with increasing the number of source nodes for fixed baud rate.

While, as seen of the curves of FQ-EDCA in figure 4.26, we remark that in FQ-EDCA, the end-to-end delay is increased slowly with increasing the number of source nodes and the baud rate of senders. That means, FQ-EDCA minimizes the end-to-end delay and enhances the QoS guarantee, as expected.

End-to-end delivery ratio

The end-to-end delivery ratio is a metric of the network performance and QoS, which can describe the percentage of received packets between sending and receiving end-points.

Thus, as seen in figure 4.27, we remark clearly that, FQ-EDCA increases the end-to-end delivery ratio better than EDCA, which enhances the QoS guarantee. While in EDCA, the end-to-end delivery ratio decreased faster than FQ-EDCA whit increasing the number of source nodes or their baud rate.

4.7 Conclusion

As expected, FQ-EDCA has proved that it can guarantee the QoS better than EDCA, with taking into account that its complexity like EDCA. Thereby, FQ-EDCA solves the most of problems of EDCA, such as fairness of resource allocations, separating the ill-behaved sources, giving the chance for the multi-hop sources to transmit, etc.

FQ-EDCA is verified for simple, large and random topologies. FQ-EDCA is evaluated comparable with EDCA. As a result, FQ-EDCA is a robust, fairly and controllable model. It has the ability for enhancing and optimizing the QoS in wireless networks.

Figure 4.26: The end-to-end delay in EDCA and FQ-EDCA for 10, 20, 30, 40 and 50 nodes

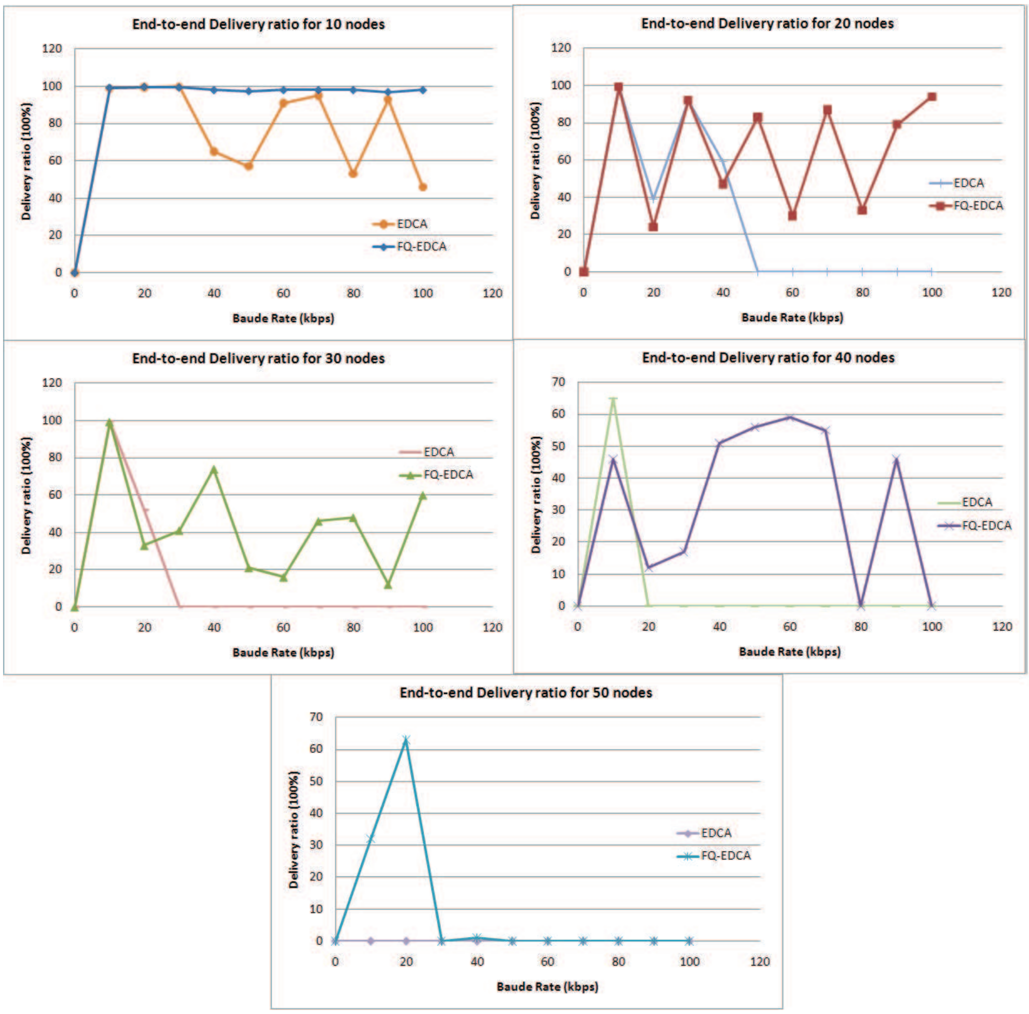


Figure 4.27: End-to-end delivery ratio in EDCA and FQ-EDCA for 10, 20, 30, 40 and 50 nodes

By the analysis of the network performance, FQ-EDCA has more performance than EDCA for throughput, delay and delivery ratio. And also, it is a powerful model for end-to-end QoS.

Chapter 5

FQ-EDCA vs. F-EDCA

In chapter 3, we have compared F-EDCA with EDCA. Thereby, it is clear by the results obtained that F-EDCA enhances EDCA by allocating the network resources with fairness. As described, F-EDCA differentiates between routing and non-routing nodes. Thus, it enhances locally EDCA by treating the packets queued in a routing node and differentiating between the traffic that belongs to that node, and the traffic that belongs to its neighbors. So, F-EDCA treats the traffic locally between the routing nodes and their neighbors.

On the other hand, in chapter 4, we have compared FQ-EDCA with EDCA. Thereby, it is clear by the results obtained that FQ-EDCA enhances EDCA by allocating the network resources with fairness. As described, FQ-EDCA differentiates the data flow packets of the source nodes to be queued in separate queues, with taking into account the control packets buffered in separate one queue. This idea enhances EDCA by supporting the ability to treat the packets of all source nodes with fairness, and separating the ill-behaved sources that consume the majority of the network resources. Therefore, FQ-EDCA enhances the end-to-end delivery packets by enhancing the end-to-end throughput and delay for all the source nodes, specially multi-hop nodes (out-ranged nodes).

However, another comparison is needed between F-EDCA and FQ-EDCA, because both of them enhance EDCA with fairness. So, this comparison will present the level of enhancement for each model and observe the different behaviors by increasing the number of source nodes and their sending rate (baud rate (kbps)). A s a result, different network performance will be obtained after the simulations of these models, as described in the next section.

5.1 The comparison of simulations

Network Simulator 2 (version 2.31 of ns-allinone (NS2)) is chosen to implement and evaluate this comparison.

Therefore multi scenarios are applied to evaluate the comparison between F-EDCA and FQ-EDCA.

However, for each scenario, the performed topology is illustrated in figure 5.1. This topology is considered as a real topology because each node is positioned randomly. This topology is generated by using topology generation tool (built on the random waypoint algorithm) implemented in the network simulator (NS2).

Also, the scenarios are performed with different baud rates (i.e. 50kbps, 100kbps and 1Mbps). However, the simulations are started by letting the sending end-point to emit its packets to the receiving end-point (destination). Then, after each two seconds, each source node starts transmitting to a random destination. For each scenario, all the source nodes are sending their packets with the same baud rate. So, each source node is sending one flow with constant baud rate, constant bit-rate CBR/UDP traffic applications and constant packet size, see table 5.1.

In addition, to realize the comparison, bottlenecks and heavy loads of traffic flows should be applied in the network. For that reason, the CBR/UDP traffic is implied in each simulation. So, this is the case considered in the performed scenarios to show the ability of F-EDCA and FQ-EDCA for treating the congestion and the queued packets.

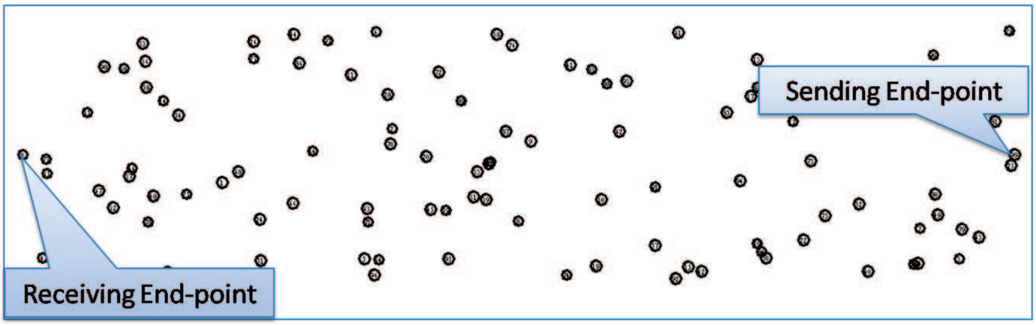


Figure 5.1: A random topology of 100 nodes

Table 5.1: The parameters of simulations

Standards	IEEE802.11b/e
QoS	EDCA
Topology mode	ad-hoc
Routing protocol	DSDV
Source nodes	1 to 99
Number of flows	1
Used ACs	VO
Type of flows	CBR/UDP
Distance between nodes	random
Transmission range	250m
Carrier sense range	550m
Topology	$2000 \times 500m$
Packet size	512 bytes
Packet baud rate	50kbps, 100kbps and 1Mbps
Bandwidth	11Mbps
CWmin	3
CWmax	15
β	0.5
Time of simulation	500s

5.2 The results of simulations

The obtained results are divided in two parts. The first one describes the network performance and the end-to-end results (i.e. end-to-end throughput and end-to-end delay) between the sending end-point to the receiving end-point for baud rates (50-100kbps). These end-points draw the longest path of routing between the sender and the receiver. On the other hand, the second part describes the network performance and the end-to-end results for high baud rate (1Mbps).

So, let us show you the evaluations of these models, as following.

5.2.1 Network performance and end-to-end results

The network performance could present the performance of a model. Thereby, it is measured by how this model exploit the network resources. Therefore, as seen in figure 5.2, we remark that the network throughput is increased with increasing the number of source nodes depending on their baud rates. Thus, the value of the network throughput approaches to the maximum value (sending baud rate) when the number of source nodes is almost 15. Then, with increasing the number of source nodes, the network throughput starts to decrease depending on the baud rate, almost for 50kbps at 50 nodes and for 100kbps at 25 nodes. As a consequence, the curve of FQ-EDCA is decreased slowly. While, the curve of F-EDCA is decreased a little bit faster than FQ-EDCA. So, this makes FQ-EDCA is better than F-EDCA for the network throughput. Further, when there are source nodes that almost cannot access the channel because of the congestion, these nodes are treated as Dead-nodes. These nodes try to attempt for accessing the channel, which is difficult to achieve, because of the congestion and the ill-behaved sources that consume the majority of network resources and make access the channel is hard. Therefore, depending on the sending baud rate, in F-EDCA, there is a lot of dead-nodes while in FQ-EDCA it is almost disappeared, because FQ-EDCA separates the ill-behaved sources.

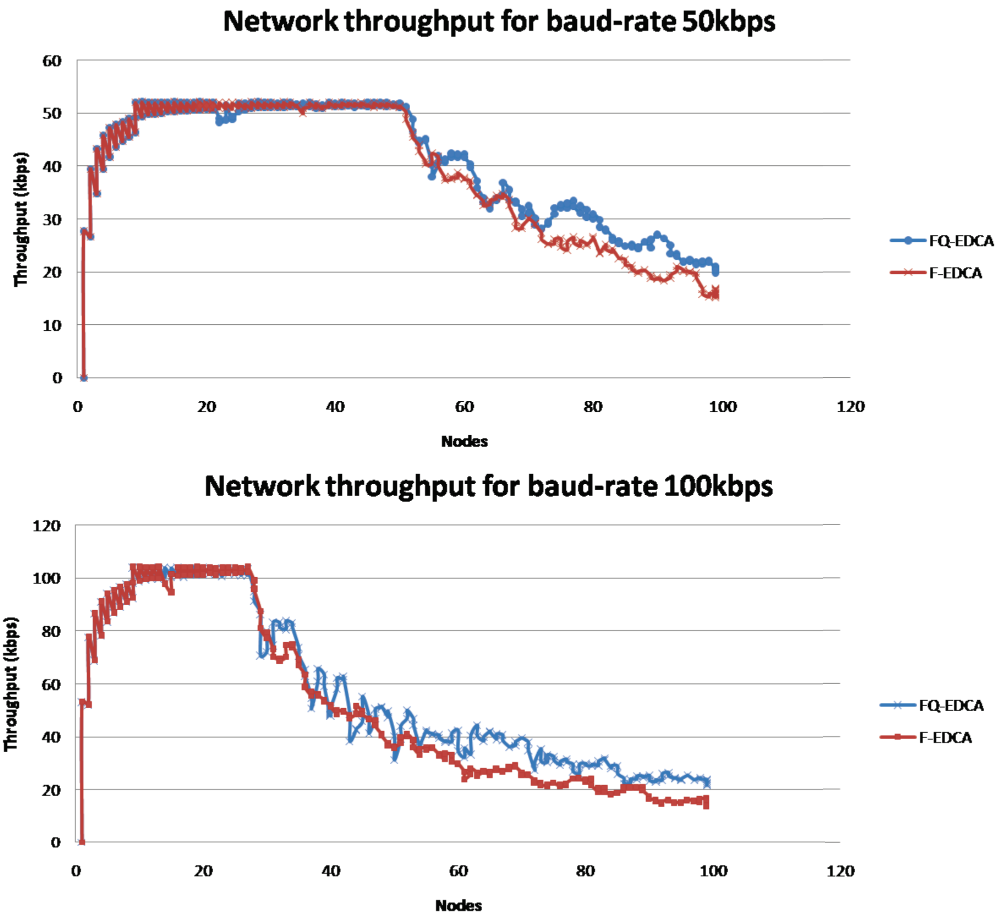


Figure 5.2: The comparison of the network throughput for different baud rates

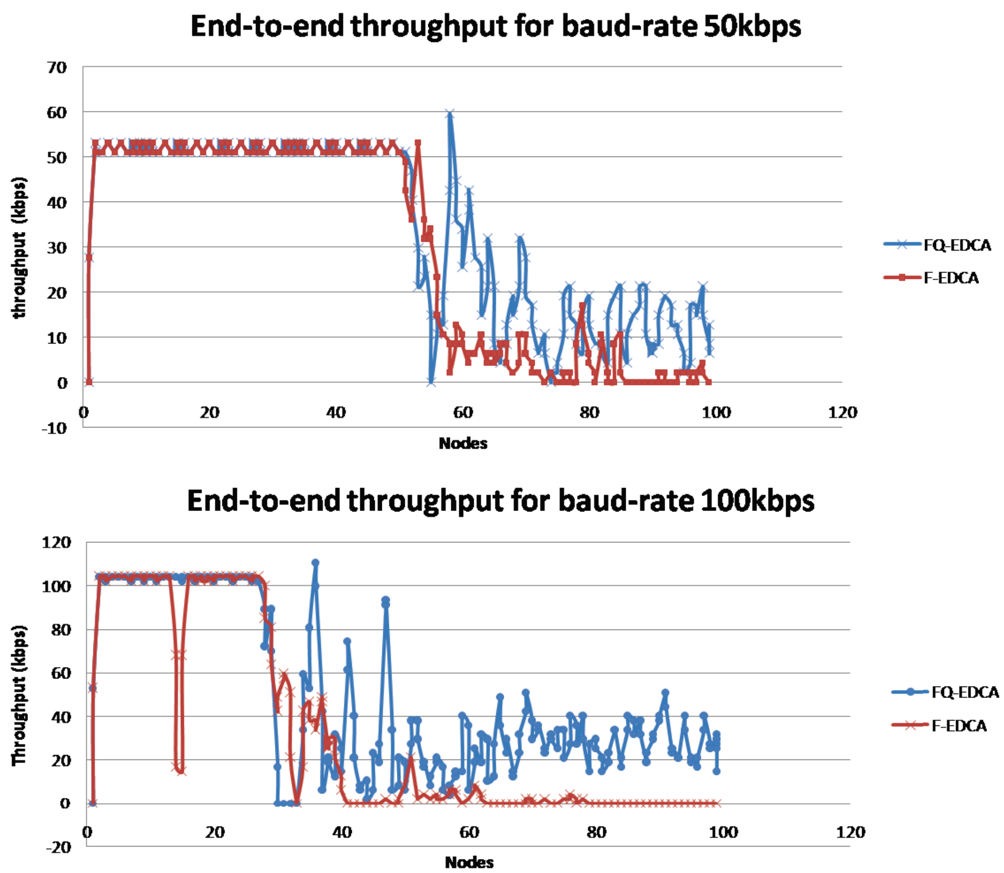


Figure 5.3: The comparison of the end-to-end throughput for different baud rates

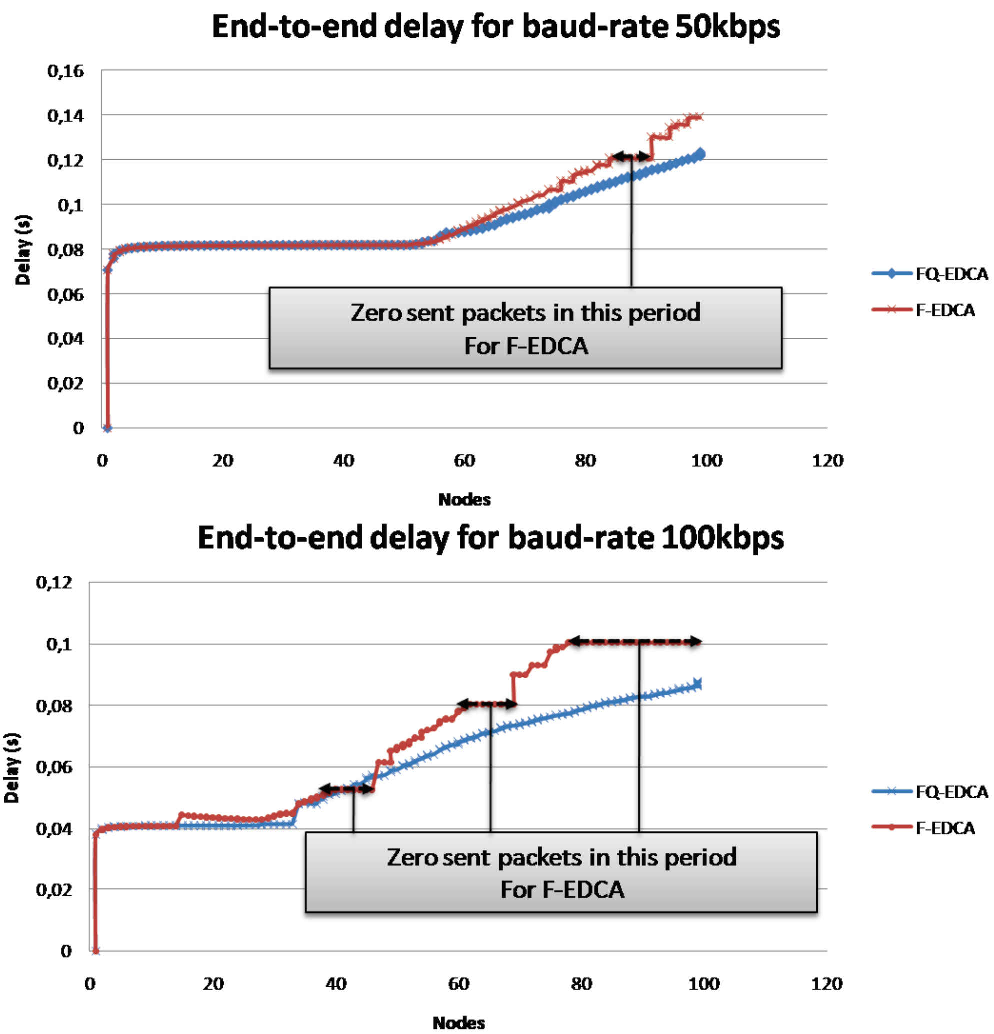


Figure 5.4: The comparison of the end-to-end delay for different baud rates

However, for the end-to-end results, figure 5.3 describes the end-to-end throughput for the longest path traffic. So, the value of the end-to-end throughput approaches to the maximum value (sending baud rate). Then, with increasing the number of source nodes, the end-to-end throughput starts to decrease depending on the baud rate, almost for 50kbps at 50 nodes and for 100kbps at 25 nodes. As a consequence for baud rate 50kbps, the average end-to-end throughput of FQ-EDCA is decreased and almost approaches to an average throughput (15kbps) for 99 source nodes. While, the average end-to-end throughput of F-EDCA is decreased and almost approaches to a low average throughput ($\approx 2kbps$) for almost 40 source nodes.

Further, as seen in figure 5.3, the average end-to-end throughput of FQ-EDCA is decreased and almost approaches to an average throughput (30kbps) for 99 source nodes. While, the average end-to-end throughput of F-EDCA is decreased and almost approaches to a low average throughput ($\approx 2kbps$) for almost 75 source nodes.

As a conclusion, FQ-EDCA is better than F-EDCA for the end-to-end throughput. Thus, the end-to-end delay is affected, as seen in figure 5.4. So, for baud rate 50kbps, the end-to-end delay started with 80ms. Then, it is increased for 50 source nodes. It is clear that the curve of F-EDCA increased faster than the curve of FQ-EDCA with taking into account that zero of packets received within the marked periods, because of the congestion and the ill-behaved sources that make access the channel is hard. Indeed, at these periods, the calculations of delay depends on the packets received. So, when there is no packets received, it consider the delay of packets is zero. That means, there is no change on the average delay calculations.

Further, as seen in figure 5.4 for baud rate 100kbps, the end-to-end delay started with 40ms. Then, it is almost increased for 15 source nodes in F-EDCA and for 25 source nodes in FQ-EDCA. It is clear that the curve of F-EDCA increased faster than the curve of FQ-EDCA with taking into account that zero of packets received within the marked periods for the problem described above.

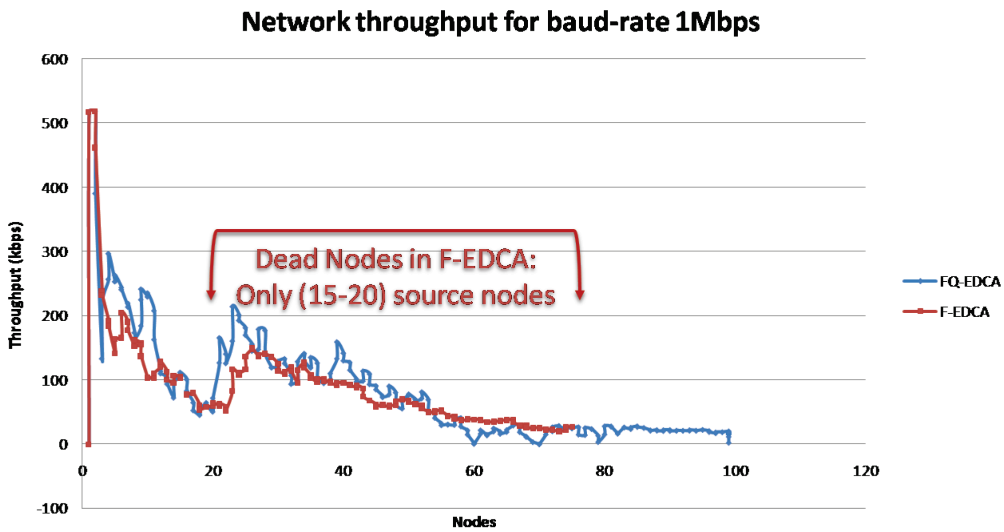


Figure 5.5: The comparison of the network throughput for baud rate 1Mbps

We remark that FQ-EDCA has better performance than F-EDCA specially for high number of source nodes. Indeed, because of the strength of FQ-EDCA, it can separate the ill-behaved sources. That leads to minimize the end-to-end delay of the longest path, as seen in figure 5.4.

5.2.2 Network performance and end-to-end results for high baud rate

Further, we will present the evaluation of the comparison between FQ-EDCA and F-EDCA when the baud rate is high (1Mbps). That means a hard traffic is applied, which leads to a hard bottleneck in the network. Therefore, as seen in figure 5.5, the network throughput in FQ-EDCA is almost like F-EDCA. Although F-EDCA still has almost the same performance of FQ-EDCA, F-EDCA does not solve the problem of the ill-behaved sources. In this simulation, they almost are 15 to 20 source nodes. Thus, the phenomena of the Dead-nodes exist in F-EDCA, because of the ill-behaved source nodes. They consume the majority of the network resources. For that reason FQ-

EDCA and F-EDCA have almost the same network throughput. On the other hand, in FQ-EDCA, there is no Dead-nodes. That means, all the nodes have the chance to send with almost the same resource allocations.

Also, with the same explication above, the end-to-end analysis for the FQ-EDCA is better than F-EDCA, as seen in figure 5.6. Where, in FQ-EDCA, the end-to-end throughput is reached almost to 40kbps as an average value when the number of source nodes is equal to 99. While, in F-EDCA, the end-to-end throughput is reached almost to zero when the number of source nodes is equal to 20.

Moreover, in figure 5.6, the end-to-end delay of F-EDCA is increased exponentially when increasing the number of source nodes. And it follows the behavior of EDCA because both of them do not separate the ill-behaved sources that consume the majority of bandwidth. While, in FQ-EDCA, the worst case of the end-to-end delay is increased almost linearly.

5.3 Conclusion

The network performance is an important metric. It almost describes the performance of how exploit the most of resources available in the network. Nevertheless, it cannot describe other existing problems like the ill-behaved sources. These sources could absorb the most of network resources without taking into account the other source nodes. Therefore, this problem leads to unfairness of resource allocation.

By the analysis of the end-to-end performance, FQ-EDCA has more performance than F-EDCA, for throughput and delay. And also, it is a powerful model for end-to-end QoS specially for a large number of source nodes. That means, in FQ-EDCA, the worst case of the end-to-end delay is increased almost linearly with increasing the number of source nodes.

As expected, FQ-EDCA enhances the QoS guarantee better than F-EDCA. Because FQ-EDCA solves the most problems in EDCA, such as fairness of resource allocations, separating the ill-behaved sources, give the

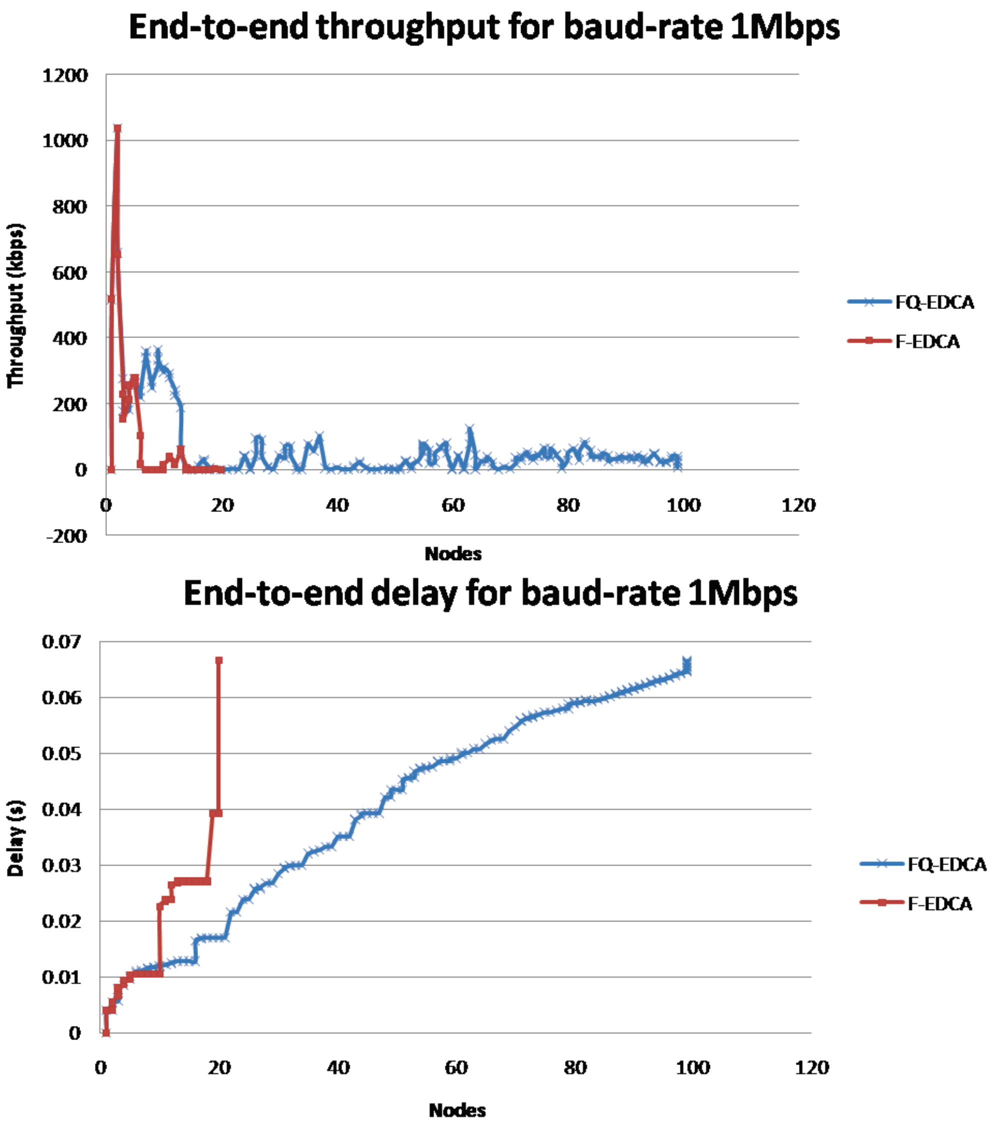


Figure 5.6: The comparison of the end-to-end throughput and delay for baud rate 1Mbps

chance to transmit for the multi-hop sources, etc. While, in the network performance (i.e. network throughput), they have almost the same behavior.

As a result, FQ-EDCA is a robust, fairly and controllable model. It has the ability for enhancing the QoS in wireless networks. While, F-EDCA is a fairly model. That has the ability to enhance locally the QoS with fairness in wireless networks.

In addition, as illustrated in FQ-EDCA curves behavior, the average network throughput is almost equal to the average end-to-end throughput. That means, the end-point source node has almost the same allocated resource as the average resource allocated to other nodes. Therefore, FQ-EDCA achieves the required fairness. Thereby, it enhances end-to-end QoS with fairness. Further, we conclude that FQ-EDCA is strong and can be applied for large topologies (as tested in 100 source nodes with high baud rate).

However, we can say that FQ-EDCA is a global model. While, F-EDCA is a local model, because F-EDCA does not solve the problem of the ill-behaved sources. So, F-EDCA treats the fairness locally between the routing nodes and their neighbors. But both of them enhanced EDCA for QoS in different judged situations. Indeed, F-EDCA treats the problem of fairness by differentiating the routing and non-routing nodes. While, FQ-EDCA treats the problem of fairness by differentiating the source nodes. Finally, we conclude that our models have good enhancement of QoS guarantees.

Chapter 6

Conclusions and perspectives

As seen in EDCA, no control management is applied to control the ill-behaved (mischievous) source nodes (i.e. In-ranged). So, the Out-ranged sources suffer from the unfairness of resource allocations. FQ-EDCA has a robust treatment of the congestion. Therefore, FQ-EDCA is classified within the queue-based enhancements. We have implemented FQ-EDCA that supports control management of traffic flows that is performed by hierarchical scheduling techniques. Thereby, it enhances EDCA by applying a fair resource allocation model to the source nodes. Thus, FQ-EDCA separates the ill-behaved sources that routing the flows of the out-ranged source nodes. As a consequence, with FQ-EDCA, the performed congestion is controlled. Each source node has a guarantee service like the others. And also, an end-to-end delay is decreased and bounded for the out-ranged sources and the average lost rate packets for all the source nodes is balanced also. So, FQ-EDCA solves most of the problems of EDCA, such as fairness of resource allocations, separating the ill-behaved sources, giving the chance to transmit for the multi-hop sources, etc.

However, the analysis of the network performance, FQ-EDCA has more performance than EDCA for throughput, delay and delivery ratio. And also, it is a powerful model for end-to-end QoS. That means, in FQ-EDCA, the worst case of the end-to-end delay is increased almost linearly when increasing the number of source nodes. While, in EDCA or F-EDCA, the end-to-end delay is increased almost exponentially. Thus, End-to-end QoS for FQ-EDCA could be applied by adopting DiffServ of network layer in wireless networks, like (Nian 04).

Further, as illustrated in FQ-EDCA results, the average network throughput is almost equal to the average end-to-end throughput. That means, the end-point source node has almost the same allocated resource as the average resource allocated to other nodes. Thereby, FQ-EDCA achieves the required fairness. Therefore, it enhances the end-to-end QoS with fairness.

On the other hand, F-EDCA is a fairness model for EDCA. Whereas, a lot of researches for QoS enhancement in ad-hoc wireless networks mainly focus on station-based DCF enhancement scheme or queue-based enhancement scheme. F-EDCA has merged the station-based DCF and queue-based enhancements. Therefore, we have implemented F-EDCA that supports resource allocation with fairness in the architecture of EDCA by taking into account differentiation between wireless node stations (Routing and Non-routing nodes). Thus, the routing node can access the channel more frequently than other nodes when there is a collision. As a consequence, each routing node has an extra bandwidth (depending on the percentage of routed packets) to compensate the deduction of its bandwidth that is needed for forwarding neighbors' packets.

Therefore, F-EDCA treats partially the hidden terminal problem and avoids the congestion by allocating fairly the resources. Thus, the misbehaving of resource allocations for non-routing nodes could be resolved locally by F-EDCA.

Further, F-EDCA could be classified as soft or local QoS approach that treats the resource allocations between a routing node and its neighbors. While, FQ-EDCA could be classified as global QoS approach that treats the resource allocations for all the source nodes.

Our future works is to combine our models (i.e. F-EDCA and FQ-EDCA) with QoS routing protocols in mobile ad-hoc networks (MANET) (IETFb).

This should enhance the fairness of allocation resources for mobile hosts (i.e. varying network topology) in ad-hoc wireless networks and guarantee the end-to-end QoS.

A perspective of this study is to evaluate FQ-EDCA, to verify the scalability of this model and to propose a model to be able to guarantee end-to-end QoS in ad-hoc networks performed for almost all the topologies.

Further, a fusion of both F-EDCA and FQ-EDCA could have a good advantage to enhance fairness for different sides, because FQ-EDCA can separate the ill-behaved sources and F-EDCA can differentiate between the routing and non-routing nodes.

However, enhancing the technology of WiMAX by adopting our models is constructive also. This enhancement could evaluate the scalability of our models to apply fairness.

Chapter 7

Annexe

7.1 Network Simulation (NS-2)

Network simulator 2 (NS2), is an object-oriented and discrete event simulator targeted at networking research (i.e. Wired, Wireless, Satellite, etc.). Thus, it is an open code application functions under Unix/Linux.

However, not all users use NS-2 under Unix/Linux (Ritc 78), because an emulator of Unix (Cygwin) (cygw) functions under Microsoft Windows (Micr).

Further, most recent release of NS-2 is ns-2.33 (released Mar 31, 2008). In our simulations, we had used NS-2.31 network simulator. That version is compatible with the EDCA model (Wiet; TKN) to be implemented.

7.1.1 Creating a wireless scenario

All configuration parameters of the simulation and the parameters of the different objects created (nodes, agents, etc.) are passed to the simulator as an input file (*.tcl).

However, there are two types of output files can be obtained as a result of simulation:

- trace files (*.tr).

- animation files (*.nam).

7.1.2 Generating tools

NS-2 has utilities to generate ad-hoc topologies and scenarios:

- setdest: generates movement pattern files using the random waypoint algorithm
- cbrgen: random traffic pattern for wireless scenarios

7.1.3 The functionality of NS-2

NS-2 defines the objects of the simulation and set their parameters. Thereby, these objects are connected to each other. Then, it defines the topology of the network. However, it starts the source applications (traffics). And at the end, it exits the simulator after a certain fixed time, or when a particular event happens.

7.1.4 Implementing a model in NS-2

To implement a model in NS-2, it is a hard work, because it depends on the functionality of the model. Indeed, there is two main languages to program the codes of the required model as:

- TCL/OTCL (TCL): Tool Command Language (TCL) for programing the events of scenarios. The object oriented of TCL (OTCL) is to code the interfaces with C++ language in NS-2.
- C/C++ (Kern 78): It is a famous object oriented language for programing the function codes of the model.

That means, if the model is an event model, it could be coded by TCL. While, C++ is the language to implement a model with functions or develop the infrastructure of model. However, in our implementation of F-EDCA and

FQ-EDCA, we had implemented EDCA, as described in (TKN). Then, we implement our models by using the C++ language in (*.cc, *.h) files. These files exist in the folder of NS-2 (ns-2.31/mac/802_11e) as following:

- F-EDCA model is implemented in the files: mac-802_11e.h, mac-802_11e.cc, priq.h, priq.cc and priority.tcl.
- FQ-EDCA model is implemented in the files: d-tail.h, d-tail.cc, mac-802_11e.h and mac-802_11e.cc.

Therefore, after implementing these models, the C++ files are ready to be compiled and used by NS-2.

Glossary

- AAS - Adaptive Antenna System
- ABR - Available Bit Rate
- ACK - Acknowledgment
- AF - Assured Forwarding
- AODV - Ad-hoc On-demand Distance Vector
- AP - Access Point
- AQR - Asynchronous QoS Routing
- ATM - Asynchronous Transfer Mode
- BA - Behavior Aggregate
- BE - Best Effort Service
- BER - Bit Error Rate
- BR - Bandwidth Routing
- BS - Base Station
- BSS - Basic Service Set

- BSSID - Basic Service Set Identification
- BWA - Broadband Wireless Access
- BW - Bandwidth
- CAC - Call Admission Control
- CBR - Constant Bit Rate
- CBWFQ - Class Based Weighted Fair Queuing
- CDMA - Code Division Multiple Access
- CEDAR - Core Extraction Distributed Ad-hoc Routing
- CFP - Contention Free Period
- COS - Class of Service
- CP - Contention Period
- CRC - Cyclic Redundancy Check
- CSMA/CA - Carrier Sense Multiple Access/Collision Avoidance
- CSMA/CD - Carrier Sense Multiple Access/Collision Detection
- CTS - Clear To Send
- DCF - Distributed Coordination Function
- DHCP - Dynamic Host Configuration Protocol
- DiffServ - Differentiated Services
- DL - Downlink
- DS - Differentiated Services
- DSCP - Differentiated Service Code Point

- DSDV - Destination Sequenced Distance Vector
- DSL - Digital Subscriber Line
- DSR - Dynamic Source Routing
- DVB - Digital Video Broadcast
- EF - Expedited Forwarding
- ESS - Extended Service Set
- FC - Fragment Control Field
- FDD - Frequency Division Duplex
- FDMA - Frequency Division Multiple Access
- FDM - Frequency Division Multiplexing
- FEC - Forward Error Correction
- FFT - Fast Fourier Transform
- FHSS - Frequency Hopping Spread Spectrum
- FTP - File Transfer Protocol
- GPS - General Process Sharing
- H-FDD - Half Frequency Division Duplex
- HSDPA - High Speed Downlink Packet Access
- HTTP - Hypertext Transfer Protocol
- IETF - Internet Engineering Task Force
- IntServ - Integrated Services
- IP - Internet Protocol

- LAN - Local Area Network
- LLC - Logical Link Control
- LLQ - Low Latency Queuing
- LMDS - Local Multichannel Distribution Service
- LoS - Line of Sight
- MAC - Medium Access Control
- MAN - Metropolitan Area Network
- MIMO - Multiple In Multiple Out
- MPDU - MAC Protocol Data Unit
- MPLS - MultiProtocol Label Switching
- MSDU - MAC Service Data Unit
- MS - Mobile Station
- NACK - Negative Acknowledgement
- NAV - Network Allocation Vector
- NodeID - Node Identifier
- OFDMA - Orthogonal Frequency Division Multiple Access
- OFDM - Orthogonal Frequency Division Multiplexing
- OLMQR - On-demand Link-state Multi-path QoS Routing
- OLSR - Optimized Link State Routing
- OQR - On-demand QoS Routing
- OSI - Open Systems Interconnection

- PAN - Personal Area Network
- PCF - Point Coordination Function
- PDU - Protocol Data Unit
- PHB - Per-Hop Behavior
- PHY - Physical Layer
- PLBQR - Predictive Location-Based QoS Routing
- QoS - Quality of Service
- QoSAODV - QoS enabled Ad-hoc On-demand Distance Vector
- R - Transmission Range
- RF - Radio Frequency
- RFCs - Request for Comments
- RREP - Route Reply
- RREQ - Route Request
- RS - Remote Station
- RSVP - Resource Reservation Protocol
- RTP - Real Time Protocol
- RTS - Request To Send
- SLA - Service Level Agreement
- SNMP - Simple Network Management Protocol
- SNR - Signal to Noise Ratio
- STA - Station

- TBP - Ticket-Based Probing
- TCP - Transmission Control Protocol
- TDMA - Time Division Multiple Access
- TDM - Time Division Multiplexing
- TDR - Trigger-based Distributed Routing
- TOS - Type Of Service
- UDP - User Datagram Protocol
- UL - Uplink
- VC - Virtual Channel
- VLAN - Virtual Local Area Network
- VoIP - Voice over Internet Protocol
- WiMax - Worldwide Interoperability for Microwave Access
- WLAN - Wireless Local Area Network
- WPAN - Wireless Personal Area Network
- WRED - Weighted Randomly Early Detected
- ZRP - Zone Routing Protocol

Bibliography

[Aad 01] I. Aad and C. Castelluccia. "Differentiation mechanisms for IEEE 802.11". *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, Vol. 1, pp. 209–218, 2001.

[Abuz 08] H. Abuzanat, B. Trouillet, and A. Toguyeni. "Routing Fairness Model for QoS Optimization in Wireless Network". *Sensor Technologies and Applications, International Conference on*, Vol. 0, pp. 776–781, 2008.

[Abuz 09a] H. Abuzanat, B. Trouillet, and A. Toguyeni. "Fair Queuing Model for EDCA to Optimize QoS in Ad-Hoc Wireless Network". *International Conference on Networking (ICN'09)*, Vol. 0, pp. 306–311, 2009.

[Abuz 09b] H. Abuzanat, B. Trouillet, and A. Toguyeni. "FQ-EDCA: An Extension of EDCA to improve Fairness in Ad-hoc Wireless Network". *International Conference on Computers & Industrial Engineering, Cie39*, pp. 1617–1622, 2009.

[Abuz 09c] H. Abuzanat, B. Trouillet, and A. Toguyeni. "Routage et équité pour améliorer la qualité de service des réseaux adhoc". *e-STA*, Vol. 6, No. 3, pp. 15–19, 2009.

[Anne 04] A. Annese, G. Boggia, P. Camarda, L. Grieco, and S. Mascolo. “Providing Delay Guarantees in IEEE 802.11e Networks”. *Vehicular Technology Conference. VTC 2004-Spring. IEEE 59th*, pp. 2234–2238, May 2004.

[Auri 04] G. Auriol, C. Chassot, and M. Diaz. “Architecture de communication à gestion automatique de la QdS en environnement IP à services différenciés”. *Technique et Science Informatiques*, Vol. 23, pp. 1137–1159, 2004.

[Bhar 94] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang. “MACAW: A Media Access Protocol for Wireless LANs”. *ACM SIGCOMM*, August 1994.

[Boda 00] S. Bodamer. “A new scheduling mechanism to provide relative differentiation for real-time IP traffic”. *In Proceeding of Global Telecommunications Conference, 2000. GLOBECOM'00*, Vol. 1, pp. 646–650, Dec 2000.

[Boui 02] E. Bouillet, D. Mitra, and K. Ramakrishnan. “The structure and management of service level agreement in networks”. *IEEE J. Select. Areas Commun*, Vol. 20, pp. 691–699, May 2002.

[Bout 07] M. Boutabia, L. Cardenas, and H. Afifi, Eds. *A Cross Layer Architecture for Realtime Applications*, Seventh International Workshop on Applications and Services in Wireless Networks, ASWN 2007, Mai 2007.

[Brad 97] R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin. “Resource ReSerVation Protocol (RSVP)”. *RFC2205, IETF*, Sept 1997.

[Brak 95] L. Brakmo and L. Peterson. “TCP Vegas: End to end congestion avoidance on a global Internet”. *IEEE J. of Selected Areas in Communication 13*, pp. 1465–1480, 1995.

[Broc 98] J. Broch, D. Maltz, D. Johnson, Y. Hu, and J. Jetcheva. "A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols". *In Proceedings of the Fourth Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom'98)*, pp. 85–97, 1998.

[Chan 07] I. Chang and S. Chen. "An End-to-End QoS Adaptation Architecture for the Integrated IntServ and DiffServ Networks". *Emerging Directions in Embedded and Ubiquitous Computing. Springer Berlin*, Vol. 4809, pp. 365–376, 2007.

[Chas 99] H. Chaskar and U. Madhow. "Fair scheduling with tunable latency: a round robin approach". *Global Telecommunications Conference, GLOBECOM '99*, Vol. 2, pp. 1328–1333, 1999.

[Chen 02] Y. Chen, Y. Tseng, J. Sheu, and P. Kuo. "On-demand, linkstate, multi-path QoS routing in a wireless mobile ad-hoc network". *In: Proceedings of European Wireless*, pp. 135–141, 2002.

[Chen 06] X. Chen, H. Zhai, X. Tian, and Y. Fang. "Supporting QoS in IEEE 802.11e wireless LANs". *IEEE Transactions on Wireless Communications*, Vol. 5, No. 8, pp. 2217–2227, August 2006.

[Chen 99] S. Chen and K. Nahrstedt. "Distributed quality-of-service routing in ad hoc networks". *IEEE Journal on Selected Areas in Communications*, Vol. 17, pp. 1488–1504, 1999.

[Cisca] i. Cisco Systems. "Class-based weighted fair queuing". http://www.cisco.com/.

[Ciscb] i. Cisco systems. "Low Latency Queuing". http://www.cisco.com/.

[Ciscc] I. Cisco Systems. "Quality of Service (QoS), Internetworking Technology Handbook". http://www.cisco.com/.

[Cisc 05] I. Cisco Systems. “DIFFSERV THE SCALABLE END-TO-END QUALITY OF SERVICE MODEL”. http://www.cisco.com/, Aug 2005. White Paper.

[Clar 98] D. Clark and W. Fang. “Explicit Allocation of Best-Effort Packet Delivery Service”. *IEEE/ACM Transactions on Networking*, Vol. 6, No. no.4, pp. 362–373, Aug 1998.

[cygw] cygwin. http://www.cygwin.com/.

[De 02] S. De, S. Das, H. Wu, and C. Qiao. “Trigger-based distributed QoS routing in mobile ad hoc networks”. *ACM SIGMOBILE Mobile Computing and Communications Review*, Vol. 6, pp. 22–35, 2002.

[Evan 07] J. Evans and C. Filsfils. *Deploying IP and MPLS QoS for Multiservice Networks: Theory and Practice*. Morgan Kaufmann, 2007.

[Fero 00] A. Feroz, S. Kalyanaraman, and A. Rao. “A TCP-Friendly Traffic Marker for IP Differentiated Services”. *Eighth International Workshop on Quality of Service (IWQoS 2000)*, pp. 138–147, Jun 2000. Pittsburgh, US.

[Floy 93] S. Floyd and V. Jacobson. “Random early detection gateways for congestion avoidance”. *IEEE/ACM Transactions on Networking*, Vol. 1, No. 4, 1993.

[Floy 94] S. Floyd. “TCP and Explicit Congestion Notification”. *ACM Computer Communication Review*, Vol. 24, No. 5, pp. 10–23, Oct 1994.

[Gall 07] J. Gallardo, P. Medina, and W. Zhuang. “QoS mechanisms for the MAC protocol of IEEE 802.11 WLANs”. *Wirel. Netw.*, Vol. 13, No. 3, pp. 335–349, 2007.

[Gamb 04] V. Gambiroza, B. Sadeghi, and E. Knightly. "End-to-end performance and fairness in multihop wireless backhaul networks". In: *Proceedings of ACM MOBICOM*, pp. 287–301, 2004.

[Gast 05] M. Gast. *802.11 wireless networks: the definitive guide*. O'Reilly, 2005.

[Gerh 03] M. Gerharz, C. Waal, M. Frank, and P. James. "A Practical View on Quality-of-Service Support in Wireless Ad Hoc Networks". *Proc. of the 3rd IEEE Workshop on Applications and Services in Wireless Networks (ASWN), Berne, Switzerland*, pp. 185–196, July 2003.

[Gole 94] S. Golestani. "A self-clocked Fair Queueing Scheme for Broadband Applications". *INFOCOM '94. Networking for Global Communications. 13th Proceedings IEEE*, Vol. 2, pp. 636–646, 1994.

[Haas 97] Z. Haas. "A New Routing Protocol For The Reconfigurable Wireless Networks". pp. 562–566, 1997.

[IEEE 05a] IEEE802.11e. "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 8: Medium Access Control (MAC) Quality of Service Enhancements". http://standards.ieee.org/getieee802/download/802.11e-2005.pdf, 2005.

[IEEE 05b] IEEE802.15.1. "IEEE Standard for Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks - Specific requirements.

Part 15.1: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Wireless Personal Area Networks (WPANs(tm))". 2005.

[IEEE 05c] IEEE802.16e. "IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands.". 2005.

[IEEE 06] IEEE802.15.4. "IEEE Standard for Information technology– Telecommunications and information exchange between systems–Local and metropolitan area networks– Specific requirements Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low Rate Wireless Personal Area Networks (LR-WPANs)". Zigbee, 2006.

[IEEE 97] IEEE802.11. "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". 1997.

[IEEE 98a] IEEE802.1D. "IEEE Standard for Information technology-Telecommunications and information exchange between systems- Local and metropolitan area networks- Common specifications Part 3: Media Access Control (MAC) Bridges". 1998.

[IEEE 98b] IEEE802.1Q. "IEEE standards for local and metropolitan area networks: virtualbridged local area networks". 1998.

[IETFa] IETF. "Internet Engineering Task Force is the body that defines standard Internet operating protocols such as TCP/IP. The

IETF is supervised by the Internet Society Internet Architecture Board (IAB). IETF members are drawn from the Internet Society's individual and organization membership".

[IETFb] IETF. "Mobile Ad-hoc Networks (MANet)". http://www.ietf.org/html.charters/manet-charter.html.

[John 96] D. Johnson and D. Maltz. "Dynamic source routing in adhoc wireless networks". *In: T. Imielinski, H. Korth (Eds.), Mobile Computing*, pp. 153–181, 1996.

[Jurd 07] R. Jurdak. *Wireless Ad Hoc and Sensor Networks: A Cross-Layer Design Perspective*. Springer, 2007.

[Kadd 07] L. Kaddour, J. Vareille, P. LeParc, and N. Berrached. "Remote control on Internet, long distance experiment of remote practice works, measurements and results". *International Review on Computers and Software (IRECOS)*, Vol. 2, No. 3, pp. 208–216, Mai 2007.

[Karn 90] P. Karn. "MACA: A New Channel Access Method for Packet Radio". *ARRL/CRRL Amateur Radio 9th Computer Networking Conference*, pp. 134–140, 1990.

[Kate 91] M. Katevenis. "Weighted Round-robin cell multiplexing in a general purpose ATM switch chip". *IEEE Journal on Selected Areas in Communications*, Vol. 9, pp. 1265–1279, Oct 1991.

[Kenn 94] A. Kennington. "Credit priority queueing for ATM switches". *ATNAC 1994 conference, Melbourne*, pp. 399–403, Dec 1994.

[Kern 78] B. Kernighan and D. Ritchie. "The C Programming Language". *Englewood Cliffs, NJ: Prentice Hall*, 1978.

[Kesh 89] S. Keshav and S. Shenker. "Analysis and Simulation of a fair Queueing Algorithm". *ACM SIGCOMM Computer Communication Review*, Vol. 19, pp. 1–12, September 1989.

[Kesh 91] S. Keshav. "A Control-Theoretic Approach to Flow Control". *Proc. ACM SigComm 1991*, Sep 1991.

[Kesh 92] S. Keshav. *Congestion Control in Computer Networks*. PhD thesis, UC Berkeley TR-654, Awarded the Sakrison Prize for the best PhD thesis in the EECS department, September 1992.

[Khan 05] H. Khanmirza, S. Zarifzadeh, and N. Yazdani. "ADPQ: An Adaptive Approach for Expedited Forwarding Traffic Scheduling". *Computers and Communications, IEEE Symposium on*, Vol. 0, pp. 801–806, 2005.

[Klei 76] L. Kleinrock. "Queuing Systems". *Computer Applications*, Vol. 2, 1976. New York: Wiley, Sons.

[Kose 08] K. Kosek, M. Natkaniec, L. Vollero, and A. Pach. "Performance Analysis of 802.11e Networks with Hidden Nodes in a Star Topology". *Consumer Communications and Networking Conference. CCNC 2008. 5th IEEE*, pp. 440–441, Jan. 2008.

[Labi 06] H. Labiod, H. Afifi, and C. Santis. *Wi-fi, Bluetooth, Zigbee and Wimax*. Springer, 2006.

[Lee 06] K. Lee, A. Toguyéni, and A. Rahmani. "Periodic Multi-Step routing algorithm for DS-TE: PEMS". *International e-Conference on Computer Science (IeCCS'2006)*, 2006.

[Lieb 99] J. Liebeherr and E. Yilmaz. "Work-conserving vs. non-workconserving packet scheduling". *Proc. IEEE/IFIP IWQoS'99*, 1999.

[Lin 01] C. Lin. “On-demand QoS routing in multihop mobile networks”. *In: Proceedings of IEEE INFOCOM 2001*, Vol. 3, pp. 1735–1744, 2001.

[Lin 99] C. Lin and J. Liu. “QoS routing in ad hoc wireless networks”. *IEEE Journal on Selected Areas in Communications*, Vol. 17, pp. 1426–1438, 1999.

[Loch 04] E. Lochin, G. Auriol, C. Chassot, A. Lozes, F. Garcia, S. Fdida, and P. Anelli. “Evaluation de la différenciation de services dans l'internet”. *Technique et Science Informatiques*, Vol. 23, No. N°5-6, pp. 675–699, 2004. Rapport LAAS.

[Lu 97] S. Lu, V. Bharghavan, and R. Srikant. “Fair scheduling in wireless packet networks”. *Proc. ACM Sigcomm, Palais des Festivals, Cannes, France*, sept 1997.

[Ma 03] L. Ma, F. Yu, V. Leung, and T. Randhawa. “A new method to support UMTS/WLAN vertical handover using SCTP”. *Vehicular Technology Conference, 2003. IEEE 58th*, Vol. 3, pp. 1788–1792, 2003.

[Maha 01] I. Mahadevan and K. Sivalingam. “Architecture and Experimental Framework for Supporting QoS in Wireless Networks Using Differentiated Services”. *Mobile Networks and Applications*, Vol. 6, No. 4, pp. 385–395, 2001.

[Mang 02] S. Mangold, S. Choi, P. May, O. Klein, G. Hiertz, and L. Stibor. “IEEE 802.11e Wireless LAN for Quality of Service”. *In European Wireless '02*, February 2002.

[Mela 02] B. Melander and M. Bjorkman. “First-come-first-served packet dispersion and implications for TCP”. *Global Telecommunications Conference. GLOBECOM'02.IEEE*, Vol. 3, pp. 2170–2174, Nov. 2002.

[Micr] Microsoft. "Window XP operating system". http://www.microsoft.com/.

[Mulr 06] P. Mulroy. "Application layer QoS for videotelephony". *BT Technology Journal*, Vol. 24, No. 2, pp. 167–173, 2006.

[Nait 03] F. Nait-Abdesselam and H. Koubaa. "RAMAC: Routing-aware Adaptive MAC in IEEE 802.11 Wireless Ad-Hoc Networks". *In 8th International Conference on Cellular and Intelligent Communications, CIC'03, Seoul, Korea*, Oct 2003.

[Nait 04] F. Nait-Abdesselam and H. Koubaa. "Enhanced Routing-Aware Adaptive MAC with Traffic Differentiation and Smoothed Contention Window in Wireless Ad-Hoc Networks". *24th International Conference on Distributed Computing Systems Workshops - W6: WWAN (ICDCSW'04)*, pp. 690–695, 2004.

[Nian 02] I. Niang, B. Zouari, H. Afifi, O. Cherkaoui, and N. Golmie, Eds. *Introduction of DiffServ QoS in 802.11 Standard*, The First Annual Mediterranean Ad Hoc Networking Workshop, Med-hoc-Net 2002, Sardegne, Italie, September 2002.

[Nian 04] I. Niang, H. Afifi, and D. Seret. "Couplage de services différenciés pour une QoS de bout en bout". *7ème Colloque Africain sur la Recherche en Informatique (CARI2004)*, 2004.

[Nich 98] K. Nichols, S. Blake, F. Baker, and D. Black. "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers". *RFC 2474, IETF*, Dec 1998.

[NS2] NS2. "Network Simulator 2". http://www.isi.edu/nsnam/ns/.

[Parc 05] P. L. Parc, J. Vareille, and L. Marcé. "Long Distance Remote Control over internet: a Reliability Challenge". *XVI Workshop on Supervising and Diagnostics of Machining Systems* *et INTELLIGENT MACHINES AND FACTORIES - Karpacz (Pologne)*, No. 1-2, pp. 222–231, March 2005.

[Pare 93] A. Parekh and R. Gallager. "A generalized processor sharing approach to flow control in integrated services networks: the singlenode case". *IEEE/ACM Transactions on Networking*, Vol. 1, No. 3, pp. 344–357, June 1993.

[Perk 00] C. Perkins, E. Royer, and S. Das. "Quality of service for ad hoc on-demand distance vector routing (work in progress)". *IETF Internet Draft, draft-ietf-manet-aodvqos-00.txt*, July 2000.

[Perk 94] C. Perkins and P. Bhagwat. "Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers". pp. 234–244, 1994.

[Perk 99] C. Perkins and E. Royer. "Ad hoc on-demand distance vector routing". *Proceedings of IEEE Workshop on Mobile Computing Systems and Applications*, pp. 90–100, 1999.

[R 96] J. R and B. Zhang. "WF2Q: Worst-case Fair Weighted Fair Queueing". *Procedings of the IEEE INFOCOM'96*, 1996.

[Rama 01] K. Ramakrishnan and S. Floyd. "The Addition of Explicit Congestion Notification (ECN) to IP". *RFC3168*, Sep 2001.

[Redd 06] T. B. Reddy, I. Karthigeyan, B. Manoj, and C. S. R. Murthy. "Quality of Service Provisioning in Ad hoc Wireless Networks: A Survey of Issues and Solutions". *Ad hoc Networks*, Vol. 4, No. 1, pp. 83–124, January 2006.

[RFC1] RFC1633. "Integrated Services in the Internet Architecture".

[RFC2a] RFC2475. "Architecture for Differentiated Service".

[RFC2b] RFC2581. "TCP Congestion Control".

[RFC2c] RFC2960. “Stream Control Transmission Protocol”.

[RFC3 03a] RFC3350. “RTP: A Transport Protocol for Real-Time Applications”. 2003.

[RFC3 03b] RFC3626. “Optimized Link State Routing Protocol (OLSR)”. 2003.

[RFC4 05] RFC4067. “Context Transfer Protocol (CXTP)”. 2005.

[RFC7 81a] RFC791. “INTERNET PROTOCOL DARPA INTERNET PROGRAM PROTOCOL”. 1981.

[RFC7 81b] RFC793. “TCP: Transmission Control Protocol”. 1981.

[Ritc 78] D. Ritchie and K. Thompson. “UNIX Time-Sharing System: The UNIX Time-Sharing System”. *B.S.T.J.*, Vol. 57, No. 6, pp. 1905–30, Aug 1978.

[Romd 03] L. Romdhani, Q. N., and T. Turletti. “Adaptive EDCF: enhanced service differentiation for IEEE 802.11 wireless ad-hoc networks”. *Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE*, Vol. 2, pp. 1373–1378, March 2003.

[Schm 04] J. Schmitt and F. Zdarsky. “A case for simplicity in providing network quality of service: class-based strict priority queuing”. *Networks, 2004. (ICON 2004). Proceedings. 12th IEEE International Conference on*, Vol. 2, pp. 809–813, Nov 2004.

[Shah 02] S. Shah and K. Nahrstedt. “Predictive Location-based QoS Routing in Mobile Ad Hoc Networks”. *Proc. IEEE ICC ’02*, pp. 1022–1027, 2002.

[Shin 03] J. Shin, C. Kuo, and D. Lee. *Quality of Service for Internet Multimedia: a General Mapping Framework. IMSC Press Multimedia Series*, Prentice Hall, 1st Ed., Jul 2003.

[Shre 96] M. Shreedhar and G. Varghese. “Efficient Fair Queuing Using Deficit Round-Robin”. *IEEE/ACM Trans. Net.*, Vol. 4, No. 4, pp. 386–397, 1996.

[Simo 03] A. Simon. “Principes de mécanismes de Qualité de Service, mesures de performances et mises en oeuvre sur Lothaire”. *Dans Actes du congrès JRES2003, Lille*, pp. 111–124, 2003.

[Sinh 99] P. Sinha, R. Sivakumar, and V. Bharghavan. “CEDAR: A Core-extraction Distributed Ad Hoc Routing Algorithm”. *IEEE JSAC*, Vol. 17, pp. 1454–1465, 1999.

[Stew 00] R. Stewart, Q. Xie, K. Morneault, C. Sharp, and V. Paxson. “Stream Control Transmission Protocol”. *RFC2960*, Oct 2000.

[TCL] TCL. “Tool Control Language”. http://www.tcl.tk/.

[TKN] TKN. “802.11e (EDCA) for NS-2”. http://www.tkn.tu-berlin.de/research/802.11e_ns2/.

[Toba 75] F. Tobagi and L. Kleinrock. “Packet Switching in Radio Channels: Part II-The Hidden Terminal Problem in Carrier Sense Multiple-Access and the Busy-Tone Solution”. *Communications, IEEE Transactions on [legacy, pre - 1988]*, Vol. 23, No. 12, pp. 1417–1433, Dec 1975.

[Tsao 00] V. Tsaoussidis and S. Wei. “QoS Management at the Transport Layer”. *ITCC 2000, Las Vegas, Nevada*, 2000.

[Vare 03] J. Vareille, P. L. Parc, L. Marcé, S. Barré, A. Mamoune, J. L. Guen, and G. Floc’h. “Contrôle actif de machines de production par Internet”. *CFM’2003 Nice (France)*, 2003.

[Vidh 03] V. Vidhyashankar, B. Manoj, and C. S. R. Murthy. “Slot allocation schemes for delay sensitive traffic support in asynchronous wireless mesh networks”. *Proceedings of HiPC 2003*, 2003.

[Wiet] S. Wiethölter and C. Hoene. "An IEEE 802.11e EDCA and CFB Simulation Model for ns-2". *Telecommunication Networks Group (TKN).* http://www.tkn.tu-berlin.de.

[Wurt 02] M. Wurtzler. *Analysis and Simulation of Weighted Random Early Detection (WRED) Queues.* PhD thesis, BSEE, Northwestern University, Evanston, Illinois, 2002.

[Xu 05] K. Xu, Q. Wang, and H. Hassanein. "Performance analysis of differentiated QoS in IEEE 802.11e WLANs". *International Journal of Communication Systems*, Vol. 18, No. 7, pp. 619–637, Mar 2005.

[Zhai 05] H. Zhai and Y. Fang. "Medium access control protocols in mobile ad hoc networks: problems and solutions". *In: Handbook on Theoretical and Algorithmic Aspects of Sensor, Ad Hoc Wireless and Peer-to-Peer Networks, edited by J. Wu, CRC Press*, pp. 231–250, 2005.

[Zhan 05] B. Zhang and H. Mouftah. "QoS Routing for Wireless Ad Hoc Networks: Problems, Algorithms, and Protocols". *IEEE Communications Magazine*, Vol. 43, No. 10, pp. 110–117, Oct 2005.

[Zhan 91] L. Zhang. "VirtualClock: a new traffic control algorithm for packet-switched networks". *ACM Trans. Comput. Syst.*, Vol. 9, No. 2, pp. 101–124, 1991.

[Zhan 96] H. Zhang. "Routing High-bandwidth Traffic in Max-min Fair Share Networks". *ACM SIGCOMM Computer Communication*, Vol. 26, pp. 206–217, Oct 1996.

[Zhon 07] S. Zhong, P. L. Parc, and J. Vareille. "Internet-based teleoperation : a case study, toward delay approximation and speed limit module". *INCINCO 2007, Angers*, Mai 2007.

Titre : Modèles d'équité pour l'amélioration de la qualité de service dans les réseaux sans fil en mode ad-hoc

L'objectif de ce travail est l'amélioration de la qualité de service (QdS) dans les réseaux sans fil ad-hoc avec équité. La QdS dans les réseaux sans fil ad-hoc est actuellement définie par la norme IEEE802.11e (EDCA). Elle permet de garantir l'accès prioritaire aux ressources pour le trafic de priorité élevé (trafic temps réel et trafic multimédia). Elle est mise en uvre dans chaque station par la classification des paquets dans différentes file d'attente caractérisant chacune une classe de trafic à laquelle est associée une priorité de traitement. Toutefois, EDCA n'est pas un protocole équitable. En effet, lorsque un nud participe au routage du trafic des ces voisins, son trafic propre se trouve réduit. Pour résoudre ce problème, nous proposons un nouveau modèle appelé F-EDCA. Ce modèle permet à un nud routeur d'accéder plus régulièrement au réseau en fonction de son taux d'occupation. Une autre forme de non équité résulte de la position d'un nud source par rapport au nud destination. Plus le nud source est éloigné, moins il a de bande passante. Pour résoudre ce problème, nous proposons FQ-EDCA. Il améliore la QdS en distinguant dans chaque classe de trafic, une file d'attente par source de trafic. Le modèle met alors en uvre des techniques d'ordonnancement équitable en se basant sur la technique du temps virtuel. Ainsi, les ressources sont allouées équitablement entre tous les nuds. F-EDCA et FQ-EDCA sont mis en uvre et évaluées de manière comparative avec EDCA. Ce travail montre que chacun d'eux améliore EDCA et pourrait allouer équitablement les ressources dans des conditions différentes et augmenter la garantie de la QdS.

Mots clés : Ad-hoc, Ordonnancement, Equité, Réseau sans fil, IEEE802.11e, EDCA.

Title: Fairness models to improve the quality of service in ad-hoc wireless networks

This thesis aims to enhance the quality of service (QoS) in wireless ad-hoc networks with fairness. The QoS in wireless ad-hoc networks is referred to the standard IEEE802.11e (EDCA) to guarantee the high priority traffic (i.e. Multimedia and Real-time traffics). Further, the QoS is managed in each station by differentiating the packets in categories to be queued depending on their priorities. Then, these packets will access the channel in different waiting time. However, EDCA cannot control the heavy traffic caused by the ill-behaved users. So, if the traffic is regulated with fairness, it could resolve some of networks' problems like the congestion and starvation. These problems degrade the QoS guarantee, because the ill-behaved sources consume the majority of the allocated resources. That leads to some of source nodes suffer from the lack of bandwidth and unfairness. So, without a proper control mechanism, it could decrease the QoS guarantees. Therefore, we propose (FQ-EDCA) that is a Fairness Queuing model for EDCA. It enhances the QoS by implementing scheduling techniques and improving the architecture of EDCA. Thereby, the traffic is regulated between the source nodes by separating the ill-behaved sources. Thus, the resources are allocated fairly between all the source nodes. Further, a fairness model (F-EDCA) is proposed to differentiate between two source nodes. One of them is routing packets of its neighbors and the other is only sending its packets. Both of these models are implemented and evaluated with EDCA. As a consequence, each of them enhances EDCA, could allocate the resources with fairness in different conditions and increase the QoS guarantee.

Keywords: Quality of Service, Ad-hoc, Scheduling, Fairness, Wireless Networks, IEEE802.11e, EDCA.

Résumé étendu

Les réseaux sans fil sont apparus dans la précédente décennie (Années 1990). Ils utilisent les ondes radio pour interconnecter différents équipements. Les techniques utilisées par les réseaux sans fil concernent principalement les transmissions dans l'infrarouge ou celles dans les bandes de fréquences libres (bandes de fréquences ne nécessitant pas de licence). Cela a permis un développement important et rapide des réseaux sans fil. En effet, l'absence de la nécessité de déploiement d'un câblage adapté a permis une utilisation flexible et peu onéreuse des solutions sans fil.
Néanmoins cette utilisation se heurte aujourd'hui à l'impact de la convergence des technologies à laquelle on assiste dans le filaire. En effet, les réseaux filaires sont aujourd'hui de plus en plus utilisés pour transporter à la fois des données numériques, mais également de la voie ou de la vidéo. Un réseau comme Internet, s'appuie sur toutes les infrastructures matérielles existantes, quelles soient filaires ou sans fil. L'objectif actuel d'Internet et d'acheminer entre utilisateurs des trafics ayant des exigences aussi différentes que les données FTP, la voix sur IP (VoIP), les flux vidéo ou les paquets de contrôle-commande des systèmes temps réels (trafic temps réel). Pour faire face à ces exigences nouvelles, il a fallu adapter Internet. C'est ainsi qu'on a introduit la notion de qualité de service (QdS) dans les réseaux en général et dans Internet en particulier.

Dans cette thèse, nous nous intéressons à garantir l'équité dans les réseaux sans fil en mode adhoc. L'équité est un moyen pour améliorer la QdS offerte à chaque nœud du réseau.

La qualité de service a pour objectif de répondre aux exigences de chaque utilisateur, de chaque application. On distingue différents critères pour mesurer la qualité de service. La bande passante et le débit sont les principaux critères utilisés. Mais on considère également souvent le délai, la gigue ou le taux de pertes en paquets émis entre une source et une destination. Dans les réseaux filaires, la QdS a été notamment mise en œuvre au niveau de la couche réseau. En effet, elle consiste dans ce cadre à mettre en œuvre des protocoles de routage dynamique qui permettent d'acheminer les différents flux en fonction de leurs exigences. Sur Internet, elle s'est donc notamment traduite par l'amélioration du protocole IP (Internet Protocol). Dans ce cadre, elle se traduit par deux grandes stratégies liées au routage IP : l'ingénierie de trafic et le routage basé sur la QdS (QoS routing).
L'ingénierie de trafic consiste à répartir les flux sur différents chemins de manière à équilibrer la répartition sur l'ensemble des réseaux d'un système autonome. Notamment, chez les fournisseurs d'accès internet (FAI), elle nécessite l'utilisation du protocole MPLS qui permet d'établir des chemins entre paires de routeurs d'entrée-sortie du domaine considéré. Les modèles développés consistent alors en des techniques de sélection de chemins répondant aux exigences de QdS d'un flux et à équilibrer ce flux sur ces chemins.
Le routage basé sur la QdS s'appuie sur la définition de files d'attente à l'intérieur des routeurs. Ces files d'attentes caractérisent des micro-flux (Intserv) ou correspondent à des agrégations de flux (Diffserv) [ACD04]. Des niveaux de priorités sont affectés à chaque file d'attente, et l'objectif de chaque routeur et de mettent en œuvre des algorithmes d'ordonnancement des paquets en tête de file d'attente afin de respecter ces priorités.

Un réseau en mode adhoc est un réseau dans lequel chaque nœud du réseau peut avoir un double rôle : d'une part être une source ou une destination d'un trafic, d'autre part participer au routage des flux qui traversent le réseau. Dans ce cadre, l'équité consiste à garantir que chaque nœud bénéficie de la même bande passante que les autres nœuds pour émettre ses flux à niveau de priorité égale. L'équité est donc un moyen d'améliorer en moyenne la QdS de

chacun des nœuds du réseau. Elle doit être garantit par rapport aux deux missions d'un nœud. On considère donc deux problématiques : l'équité de routage et l'équité de source.

Dans les réseaux sans fil en mode adhoc, le mécanisme de base utilisé par les nœuds dans le cadre de la norme IEEE 802.11 est le DCF (Distributed Coordination Function). Ce mécanisme ne considère par la qualité de service. Il a été amélioré dans le cadre de la norme IEEE 802.11e, au travers notamment du modèle EDCA (Enhanced Distributed Channel Access) [IEE05a], [CZTF06]. La mise en œuvre de la QdS dans les réseaux sans fil est basée sur le constat que les modèles des réseaux filaires comme Intserv et Diffserv sont inadaptés en raison de la problématique spécifique du sans fil. Le caractère spécifique de ces réseaux provient notamment de la méthode d'accès qui entraîne une compétition entre les nœuds pour accéder au support de communication partagé en cas de besoin simultanée. On n'est pas comme dans le filaire ou sur des liaisons en point à point, le seul soucis qu'on a est de définir parmi toutes la files d'attentes internes à un nœud, laquelle sera la prochaine à être servie. Aussi, l'idée d'EDCA a consisté à s'inspirer au niveau de la couche MAC de ce qui a été fait en filaire au niveau de la couche réseau : définir différentes files d'attente en fonction de la classe de trafic. EDCA définit ainsi 4 classes de trafic de priorité décroissante : AC_Vo (Voix et trafic temps réel), AC_Vi (vidéo), AC_Be (Best effort – Autre trafic TCP), AC_Bk (Le reste comme le trafic UDP). A ces différentes classes de trafic sont associées différents priorités qui se traduisent implicitement, d'une part par des temps d'attente différenciés pour l'accès au réseau comme défini par l'équation (1) :

AIFS(AC) = slot-time*AIFSN(AC)+ SIFS (1)
avec AC la classe de trafic considérée.

	CW min	CW max	AIFSN
AC_Vo	7	15	2
AC_Vi	15	31	2
AC_Be	31	1023	3
AC_Bk	31	1023	7

Figure 1. Paramètres associés aux différentes classes de EDCA

Et d'autre part, en cas de collision un temps d'attente plus important pour le trafic des classes les moins prioritaires comme spécifié par les équations (2) et (3)

Backoff(AC) = rand(0,CW(AC)) (2)
CW(AC) = min {2*CW(AC), CWmax(AC)} (3)

Malgré l'avancée que représente un modèle comme EDCA, nous avons montré que ce n'est pas un modèle équitable du point de vue accès au support de transmission.

Le premier problème concerne l'équité de routage. L'équité de routage consiste à faire en sorte qu'un nœud qui route les trafics émis par ses voisins ne soit pas pénalisé par cette mission par rapport à son flux propre. Considérons par exemple l'exemple donné par la Figure 2.

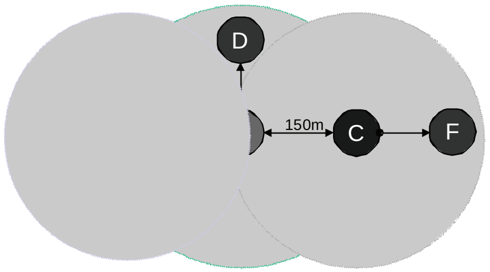


Figure 2. Exemple illustratif pour l'évaluation de l'équité de routage

Dans cet exemple, nous considérons que les trois nœuds nommés A, B, et C envoient des flux vers différentes destinations (D, E et F). Certains flux de A et C sont routés par le nœud B.

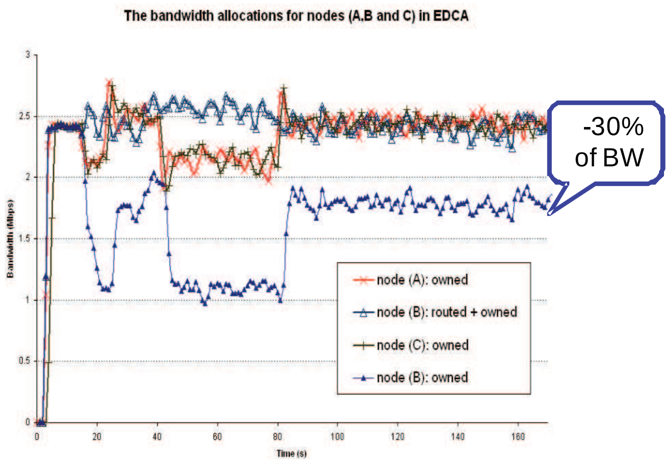


Figure 3. Mesure de la bande passante allouée à chaque nœud de la topologie Figure 2.

La Figure 3 nous montre que le nœud B reçoit la même bande passante que A et C pour sont flux propre et les flux qu'il route. En conséquence, il a moins de bande passante pour son flux propre que les autres nœuds (cf. courbe bleue sur Figure 3). C'est donc une illustration de l'iniquité de EDCA.

Pour y remédier, nous proposons le modèle F-EDCA (Fairness EDCA) [ATT08]. Ce modèle est une amélioration de EDCA. Il consiste à définir deux files d'attente dans chaque classe d'accès : une file d'attente pour le flux propre et une file d'attente pour le flux routé. Ensuite pour chaque classe, la fenêtre de contention utilisée pour le calcul du backoff est définie par les équations (4), (5) et (6) :

$$CW_{ACi}new = min(MF^e_{ACi} * CW_{ACi}old, CW_{ACi}max)$$

(4)

$$MF^e_{ACi} = 1 + \beta_{AC} * \rho^e_{ACi}$$

(5)

$$\rho^e_{ACi}(t) = Wo_{ACi}(t)/(Wo_{ACi}(t) + Wr_{ACi}(t))$$

(6)

L'idée ici est qu'en cas de collision le CWnew n'est pas obtenu par un simple doublement de la fenêtre. Il dépend d'abord de MF^e_{ACi}, qui est un facteur multiplicateur qui est compris entre les valeurs 1 et 2. Plus le flux routé est grand, plus ce facteur est proche de 1(cf. équation 6 qui définit le paramètre ρ^e_{ACi} qui est inversement proportionnel au flux routé Wr_{ACi}). Cela signifie que plus un nœud route un flux, plus il sera favorisé pour l'accès au canal. Dans cette équation, nous définissons un paramètre β_{AC}, qui dépend de la classe de trafic. Ce paramètre est compris entre 0 et 1 et est d'autant plus petit que la classe est prioritaire. Cela permet de renforcer le caractère prioritaire d'un trafic en rendant son facteur multiplificateur plus proche de 1.

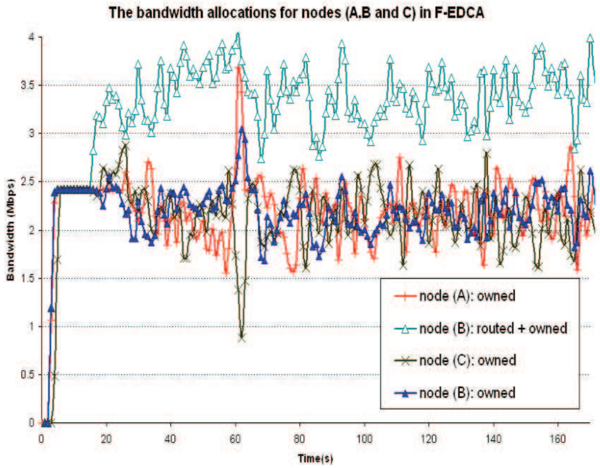


Figure 4. Partage de la bande passante entre les nœuds A, B et C dans le cas de F-EDCA

L'implémentation et la simulation de ce modèle sous NS2 nous a permis de montrer qu'il résout l'équité de routage (Figure 4). La Figure 4 montre que du fait que le nœud B est également une fonction de routage, le modèle F-EDCA lui permet d'accéder plus souvent au médium que les deux autres nœuds avec lesquels il est en compétition (cf. courbe en bleu ciel). La bande passante qui est allouée à son flux propre est du même ordre que celle des nœuds A et C (cf. courbe en bleu foncé).
Les évaluations réalisées avec NS2 nous ont également permis de montrer que F-EDCA permet d'améliorer les délais aussi bien du flux propre de B mais également des flux qu'il route.

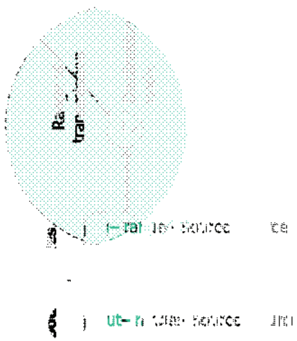


Figure 5. Exemple illustratif de la problématique de l'équité de source

La deuxième problématique que nous traitons dans ce travail de thèse est l'équité de source. Cette problématique est illustrée par la Figure 5. Elle se résume au fait que lorsqu'on considère deux nœuds sources émettant vers une même destination, plus un nœud est éloignée de la destination, moins ses flux ont de la bande passante pour atteindre la destination. Comparons par exemple, les situations des nœuds A et B de la Figure 5. Le nœud A se situe trop loin de C (nœud « out-ranged source ») pour pouvoir lui transmettre directement ces donnés. Il doit les faire transiter par B. Le nœud B lui par contre communique directement avec C car il la distance qui les sépare le permet (nœud « in-ranged source »). L'iniquité provient du fait que le nœud B va accorder moins de bande passante aux flux de A qu'à ces flux. Ainsi, plus un nœud sera éloigné d'une destination, plus ses flux seront pénalisés.

Pour faire résoudre ce problème, nous proposons le modèle FQ-EDCA (Fairness Queuing – EDCA) [ATT09a], [ATT09b]. Ce modèle est de nouveau une extension de EDCA. Il considère donc les 4 classes d'accès d'EDCA. Dans chacune de ces classes on va définir une file d'attente par nœud source, et une file d'attente pour les paquets de contrôle.

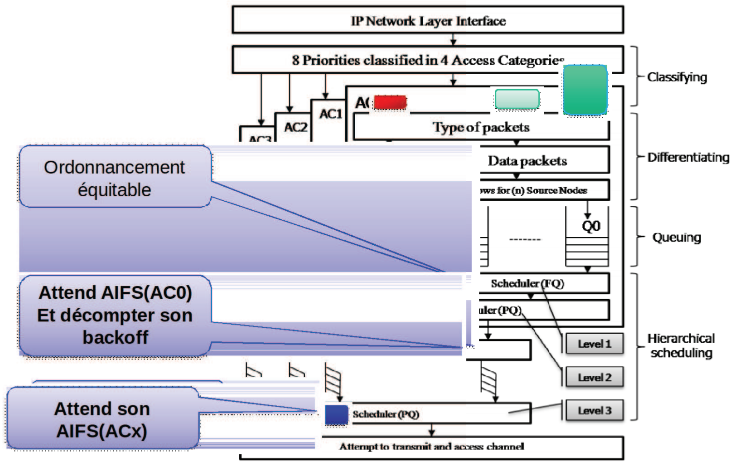


Figure 6. Ordonnancement hiérachique de FQ-EDCA

La Figure 6, schématise le principe de ce modèle à l'intérieur d'un nœud. L'idée est de faire en sorte que toutes les sources soient servies se voient allouer la même quantité de bande passante, à l'intérieur d'une des 4 catégories de flux. Pour cela, nous avons défini un ordonnancement équitable qui procède à la sélection d'un paquet en tête des files d'attente au sein d'une catégorie. Cet ordonnancement équitable (Fair Queing Scheduling) est basé sur le modèle mathématique défini par les équations (7) à (11). L'idée directrice de ce modèle est la définition d'une horloge virtuelle notés 'V' qui permet de voir si un flux est en avance ou retard d'un point de vue traitement. La priorité sera donnée au traitement des flux en retard.

Notations :
P^{q}_{kq} est le qème paquet de la file d'attente k ;
V est l'horloge virtuelle ;

$L(P^q_{kq})$ est la longueur en octets du paquet P^q_{kq} ;

$S(P^q_{kq})$ est la date potentielle de démarrage du service du paquet ;

$F(P^q_{kq})$ est la date potentielle de fin de service du paquet ;

n_{aci} est le nombre de files d'attentes dans la classe considérée du nœud i.

$$S(P^q_{kq}) = \max(V, F(P^q_{kq-1}))$$
(7)

L'équation (7) permet de calculer la date potentielle de démarrage du paquet en tête de chaque file d'attente k au sein de la classe de trafic.

$$F(P^q_{kq}) = S(P^q_{kq}) + L(P^q_{kq}) \qquad (8)$$

L'équation (8) sert au calcul de la date de fin de service du paquet en tête de la file d'attente k, compte tenu de sa longueur.

$$S_{\min} = \min_{1 \le q \le n_{aci}} \left\{ \; S(P^p_{kq}) \; \right\}$$
(9)

L'équation (9) permet de calculer la date de service au plus tôt parmi toutes les dates associées aux différentes files d'attentes.

$$P_e = \arg\min_{\{kq / S(p^q_{kq}) \le V\}} \left\{ \; F(P^p_{kq}) \; \right\} \qquad (10)$$

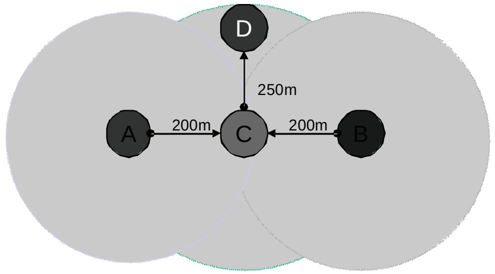
L'équation (10) permet d'identifier la file d'attente traitée. C'est parmi les files d'attentes en retards ($S(P^q_{kq}) <= V$) celle qui serait la première à se terminer. Cela permet de privilégier les files d'attentes contenant des petits paquets comme au détriment de celles qui ont des paquets de taille importante.

$$V = \max(V_{old} + \frac{L(P_e)}{n_{aci}}, S_{\min})$$
(11)

L'équation (11) permet après chaque cycle de traitement recalculer la date de l'horloge virtuelle V.

Une fois sélectionné par le mécanisme de FQ-scheduling, le paquet va passer deux autres ordonnancements basés sur la priorité comparée des flux de manière à résoudre les collisions virtuelles avec soit un paquet de contrôle, voire des paquets sélectionnés pour les autres classes de trafic.

Ce modèle a était évaluer par simulation sous NS2 de manière d'une part à vérifier sa capacité à résoudre le problème d'équité de source. Pour cela nous l'avons comparé à EDCA sur des topologies facilement analysables comme celle de la Figure 7.



La norme	IEEE802.11b/e
QoS	EDCA
Protocole de routage	DSDV
Noeud puits	D
Les noeuds source	A,B and C
Le nombre de flux	4 (1 flux pour chaque AC)
Le type de flux	CBR/UDP
Transmission range	250m
Carrier sense range	550m
Le taille de paquet	1472 octets
Le débit de paquet	1Mbps
Bande passante	11Mbps
Le temps de simulation	200s

Figure 7. Exemple de topologie pour évaluer l'équité de flux de FQ-EDCA

Figure 8. Illustration de l'iniquité de flux pour EDCA

Les Figure 8 et Figure 9 illustrent la différence d'équité entre EDCA et FQ-EDCA. On peut voir que pour toutes les classes de trafic, après un transitoire de quelques ms, toutes les sources se voient allouer la même quantité de bande passante. FQ-EDCA résout donc bien l'équité de flux.

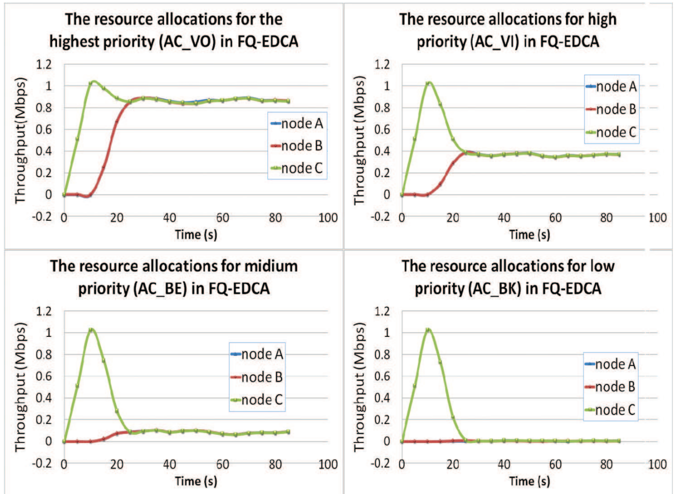


Figure 9. Illustration de l'équité de flux pour FQ-EDCA

A l'occasion de ce travail, nous avons fait d'autres évaluations ayant notamment pour objectif de vérifier le passage à l'échelle des deux modèles proposés dans cette étude. Nous avons montrés que ces deux modèles ont une complexité du même ordre que celle de EDCA. Donc nous pouvons penser qu'il devrait résister au facteur d'échelle, même si cela mérite d'être approfondit dans le cas de FQ-EDCA qui présente des analogies avec le modèle Intserv qui est connu pour ne pas avoir résisté au facteur d'échelle. Nous avons montré sur des simulations à plus grande échelle, que FQ-EDCA semble d'une manière générale, plus équitable que F-EDCA. Mais cela doit être confirmé. D'une manière générale, nous pensons qu'il est possible de fusionner ces deux modèles pour avoir un modèle qui résout les deux types d'iniquités que nous avons traité dans cette étude.

Une autre perspective de ce travail réside dans la définition formelle du paramètre β que nous avons introduit dans le modèle de F-EDCA. Notre souhait est de pas le définir de manière empirique, mais de tenir compte de facteurs liés aux caractéristiques du flux et à l'environnement du nœud (connexité …) pour calculer dynamiquement ce paramètre.

Parmi les autres perspectives de ce travail, nous pouvons penser à faire le lien entre le traitement différentié des flux fait dans la sous-couche MAC qui améliore localement la QdS et la problématique de la QdS de bout-en-bout entre un émetteur et un destinataire [GSK04], [IC07].