THESE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par *l'Université Toulouse III - Paul Sabatier*
Discipline ou spécialité : *Informatique*

Présentée et soutenue par *BOUDOUR Ghalem*
Le *24 Septembre 2010*

Titre : *Bandwidth reservation in mobile ad hoc networks for providing QoS – Adaptation for voice support*

JURY
Professeur Abderrahim Benslimane, Université d'Avignon - Rapporteur
Professeur Congduc Pham, Université de Pau - Rapporteur
Professeur Guy Juanole, Université Toulouse III - Examinateur
Professeur Thierry Val, Université de Toulouse II - Examinateur
Professeur Francine Krief, ENSEIRB de Bordeaux - Examinatrice
Professeur Mammeri Zoubir, Université Toulouse III - Directeur de thèse

---

Ecole doctorale : *Mathématiques, Informatique et Télécommunications de Toulouse*
Unité de recherche : *Institut de Recherche en Informatique de Toulouse – UMR 5505*
Directeur(s) de Thèse : *Mammeri Zoubir*
Rapporteurs : *Abderrahim Benslimane* *Professeur, Université d'Avignon*
Congduc Pham *Professeur, Université de Pau*

In the Name of Allah, the Benificent, the Merciful

Au nom d'Allah, le Très Miséricordieux, le Tout Miséricordieux

"My prayer and my sacrifice and my life and my death are surely for Allah, the Lord of the worlds"

Le Saint Coran

Exclusivement à mon père pour ses
énormes sacrifices

Remerciements

Les remerciements ! Que c'étaient dur le travail avant d'arriver à écrire cette partie du mémoire. J'imagine que ça aurait été plus dur sans le soutien de plusieurs personnes que je tiens à leur exprimer ma reconnaissance et gratitude dans quelques lignes.

Tout d'abord je tiens à remercier chaleureusement les membres du jury. Merci au Professeur Abderrahim Benslimane (Université d'Avignon), et au Professeur Congduc Pham (Université de Pau) pour avoir accepter de rapporter mon travail. Mes remerciements s'adressent aussi au Professeur Guy Juanole (LAAS-CNRS), au Professeur Thierry Val (Université de Toulouse), et à Madame la professeur Francine Krief (ENSEIRB de Bordeaux), qui ont fait l'honneur de participer au jury de ma soutenance de thèse.

Je tiens à remercier particulièrement mon directeur de thèse le Professeur Zoubir Mammeri, qui était toujours présente à mes coté pendant la réalisation de ce travail. Ses conseils, orientations, remarques, et soutien morale ont été d'une importance crucial dans l'aboutissement de ce travail. Un grand merci chef, pour ta patience, et pour le temps que tu m'as consacré durant ces quatre années difficiles.

Un grand Merci à Cédric teyssié pour son soutient, ses conseils techniques, ses discussions intéressantes, ainsi pour ses orientations au niveau de l'enseignement. Je tiens à remercier l'ensemble des membres de mon équipe, et particulièrement mes collègues avec lesquels j'ai partagé mon bureau: sacré Jo et Wassim, ainsi que les autres qui sont dans l'autre bou de l'IRIT, je veux dire par cela Mahboub, Imran, et Mehdi, sans oublier le nouveau arrivant, Rahim à qui je souhaite beaucoup de succès et de réussite dans cette équipe.

Je ne peux pas mettre fin à cette partie de la thèse avant de remercier mes amis de Toulouse avec lesquels je partage mon quotidien (Khallil, Nadim, Ibrahim, Houssem, Alaa, Saâd, Houssem Jerbi, Nacim, Yassine, Sami, …), ainsi que mes amis du bléd (nounou, Boumediene, Hamid, Abdou, Gézza, Habib, …).

Finalement, je veux remercier les personnes sans lesquelles mon existance et tout ce travail n'auraient aucun sens. Je remercie profondément mon père, ma mère, mon frère, et mes sœurs qui m'ont soutenu par tous les moyens, je serai reconnaissant à vie et je vous dis que je vous aime beaucoup.

Résumé

Le support de qualité de service (QoS) dans les réseaux MANETs (Mobile Ad-Hoc NETworks) a attiré une grande attention ces dernières années. Bien que beaucoup de travaux de recherche ont été consacré pour offrir la QoS dans les réseaux filaires et cellulaires, les solutions de QoS pour le support du trafic temps réel dans les MANET reste l'un des domaines de recherche les plus difficiles et les moins explorés. En fait, les applications temps réel telles que la voix et la vidéo ne pourrait pas fonctionner correctement dans les MANET sans l'utilisation d'un protocole de contrôle d'accès au support (MAC) orienté QoS. En effet, les trafics temps réel demandent des exigences strictes en termes de délai de transmission et de taux de perte de paquets qui peuvent être remplies uniquement si la sous-couche MAC fournit un délai d'accès au canal borné, et un faible taux de collision.

Le but de cette thèse est la proposition et l'analyse d'un protocole MAC basé sur la réservation pour garantir la QoS dans les MANETs. Tout d'abord, nous étudions un problème majeur dans la réservation de ressources dans les MANETs qui est la cohérence des réservations. Notre analyse des protocoles de réservation existant pour les MANETs révèle que de nombreux conflits de réservations entre les nœuds voisins se produisent pendant la phase d'établissement de réservation. Ces conflits, qui sont principalement dues à la collision des messages de contrôle de réservation, ont un impact important sur les performances du protocole de réservation, et conduisent à un taux de collision et de perte de paquet importants pendant la durée de vie de la connexion, ce qui n'est pas acceptable pour les trafics temps réels. Nous proposons un nouveau protocole MAC basé sur la réservation qui résout ces conflits. Le principe de notre protocole est d'établir une meilleure coordination entre les nœuds voisins afin d'assurer la cohérence des réservations. Ainsi, avant de considérer qu'une réservation est réussite, le protocole s'assure que chaque message de contrôle envoyé par un nœud pour établir une réservation est bien reçu par tous ses nœuds voisins.

Dans la deuxième partie de cette thèse, nous appliquons le protocole de réservation proposé au trafic de type voix. Ainsi, nous étendons ce protocole afin de prendre en compte les caractéristiques du trafic voix, tout en permettant le transport de trafic de données. Nous nous focalisons sur l'utilisation efficace de la bande passante et les mécanismes pour réduire le gaspillage de bande passante.

La dernière partie de cette thèse concerne l'extension du protocole proposé en vue de réserver la bande passante pour une connexion temps réel sur un chemin. Ainsi, le protocole MAC de réservation proposé est couplé avec un protocole de routage réactif. En outre, le protocole est étendu avec des mécanismes de gestion de à mobilité afin de faire face à la dégradation des performances due à la mobilité des nœuds.

Nous évaluons les performances du protocole proposé dans plusieurs scénarios dans lesquels nous montrons sa supériorité par rapport aux standards existants.

Mots clefs: Réseaux sans fil Ad-hoc, Qualité de Service, réservation de bande passante, trafic voix, protocoles MAC.

Abstract

QoS provisioning over Mobile Ad-Hoc Networks (MANETs) has attracted a great attention in recent years. While much research effort has been devoted to provide QoS over wired and cellular networks, QoS solutions for the support of real-time traffic over MANETs remains one of the most challenging and least explored areas. In fact, real-time applications such as voice and video could not function properly on MANETs without a QoS oriented medium access control (MAC) scheme. Indeed, real-time traffics claim strict requirements in terms of transmission delay and packet dropping that can be fulfilled only if the MAC sub-layer provides bounded channel access delay, and low collision rate.

The purpose of this thesis is the proposal and analysis of an efficient reservation MAC protocol to provide QoS support over MANETs. Firstly, we study one major issue in resource reservation for MANETs which is reservation consistency. Our analysis of existing reservation MAC protocols for MANETs reveals that many reservation conflicts between neighbor nodes occur during the reservation establishment phase. These conflicts which are mainly due to collisions of reservation control messages, have an important impact on the performance of the reservation protocol, and lead to a significant collision and loss of packets during the life-time of the connection, which is not acceptable for real-time traffics. We design a new reservation MAC protocol that resolves these conflicts. The main principle of our protocol is to achieve better coordination between neighbor nodes in order to ensure consistency of reservations. Thus, before considering a reservation as successful, the protocol tries to ensure that each reservation control message transmitted by a node is successfully received by all its neighbors.

In the second part of this thesis, we apply the proposed reservation protocol to voice traffic. Thus, we extend this protocol in order to take into account the characteristics of voice traffic, while enabling data traffic. We focus on efficient bandwidth utilization and mechanisms to reduce the waste of bandwidth.

The last part of this thesis relates to the extension of the proposed protocol in order to reserve resources for a real-time connection along a path. Thus, the proposed reservation MAC protocol is coupled with a reactive routing protocol. In addition, the protocol is extended with mobility handling mechanisms in order to cope with performance degradation due to mobility of nodes.

We evaluate the performance of the proposed scheme in several scenarios where we show its superiority compared to existing standards.

Keywords: Wireless ad-hoc networks, Quality of Service, Bandwidth reservation, Voice traffic, MAC protocols.

Table of Contents

Table of Figures

Table of Tables

Introduction

Context

Wired Local Area Networks had been for long time the primary means of communication. However, this technology does not longer fulfill mobility requirements of end-users. Thus, with the emergence of low-cost wireless equipments (such as laptops, PDAs, and mobile phones) end-users want to move freely while keeping connected to their personal or enterprise network. A solution to offer wireless connectivity to end-users consists in using an infrastructure-based wireless network architecture where communications are controlled by an access point (AP). Despite this architecture allows a flexible mobility to end-users it can be used only in restricted areas such as enterprises or campus. In addition, the fact that this architecture is centralized makes it non-fault tolerant if the AP crashes. Due to these reasons, an infrastructure-based network is not always possible. By consequence, alternative wireless network architectures that do not rely on any infrastructure are more than a requirement. Among such infrastructure-less network architectures, we have wireless ad-hoc networks which are expected to play an important role in future generation networks.

A Wireless ad-hoc network is a wireless network able to self-organize without previously defined infrastructure. Such a network consists in mobile stations or nodes that can communicate directly with each other if they are located within transmission range of each other. The transmission range of mobile stations is relatively limited. Consequently, the deployment of a large-scale network requires that the ad-hoc network is multi-hop, i.e. intermediate stations act as relay nodes. Due to their self-organization and lack of infrastructure, wireless ad-hoc networks can be easily deployed in many areas such as embedded (built recently in vehicles to increase the safety of users informing them of any obstacles on their route), during rescue operations in disaster areas, or in military operations. Such networks are also characterized by their limited resources such as limited batteries, and processing capacity, leading to limited autonomy. Moreover, the capacity of wireless links is relatively limited leading to low throughput in comparison with wired networks.

Wireless ad-hoc network users want to have the same services as those offered by wired networks. In other words, the applications used in wired networks should be available on ad-hoc networks, in particular, multimedia and real-time applications (such as video conferencing, Internet telephony, video on demand ...). In order to be useful for the end-users, the data transmitted by these applications should meet some Quality of Service (QoS) requirements such as minimum guaranteed bandwidth, maximum transmission delay, and maximum dropping rate. However, these QoS requirements can be delivered to the end-users only if an efficient medium access control (MAC) scheme is used at the data link layer.

MAC protocols control access to the wireless medium, and define how mobile nodes can share the limited wireless bandwidth resource in an efficient manner. They also are in charge of resolving conflicts among contending nodes for channel access. Thus, they have a significant impact on the QoS provided to real-time traffics in terms of transmission delay, delay jitter, and dropping rate. QoS requirements can be provided only if MAC protocols grant channel to nodes in such a manner that increases bandwidth utilization, and minimizes collisions and channel access delay.

Due to the variable nature of the wireless channel, nodes mobility, and limited bandwidth in wireless ad-hoc networks, the design of MAC protocols that provide QoS guarantees over such networks is very challenging. The popular IEEE 802.11 DCF scheme does not include any explicit QoS support. All packets are treated and served in exactly the same manner. Thus, the packet delay, the dropping rate, and the offered session throughput cannot be predicted because they depend on the network conditions such as traffic load and node mobility. The lack of predictability and controllability largely limit the DCF scheme in ad hoc networks to best effort applications only.

In recent years, numerous research efforts have been devoted to devise new MAC schemes, which are called QoS-aware MAC protocols. The first category of these protocols consists in making QoS extensions to the legacy DCF scheme through service differentiation. Prioritization MAC protocols (such as IEEE 802.11e [IEE 05]) allow mobile terminals to access the wireless medium in a differentiated manner, in order to satisfy the QoS requirements of high priority flows. The IEEE 802.11e EDCF standard is the most representative of these protocols.

Despite service differentiation enhances significantly the performance of the legacy IEEE 802.11, most of prioritization schemes brought to the IEEE 802.11 provide only a *soft* QoS. Indeed, in order to provide *deterministic* QoS, an allocation-based approach is necessary.

Reservation-based MAC protocols aim at providing QoS guarantees through ensuring a deterministic access to the wireless channel. These protocols include some features which are interesting for real-time and multimedia traffics such as low collision rate, low access delay, and low impact of traffic load. Thus, our research work in this thesis is oriented toward proposing an efficient reservation-based MAC protocol for QoS provisioning in wireless ad-hoc networks.

Contributions

The first contribution consists in proposing a reservation-based MAC protocol achieving efficient bandwidth utilization through reducing the collision rate. The ability of a reservation MAC protocol to provide a collision-free schedule depends on the ability of nodes to inform their neighbors about their current reservations, that is, to prevent neighbor nodes from reserving already reserved slots and ensure consistency of reservations. After an analysis of existing reservation-based MAC protocols, we found that most of these protocols suffer inconsistency of reservations. Reservation inconsistency occurs when some conflicts of reservations appear between neighbor nodes, because some of these nodes are not aware of reservations established by their neighbors. Such nodes may try to reserve already reserved slots causing loss of reservation and collision during reserved slots. After highlighting the various factors (which are mainly collisions of reservation control packets) involved in the occurrence of such reservation conflicts, we propose a reservation protocol that allows nodes to establish consistent reservations while avoiding such conflicts through better coordination and cooperation between nodes. Indeed, the handshake scheme of our protocol consists in ensuring that a reservation is confirmed and considered only if it is recorded by all the neighbors of both the sender and receiver. Thus, any unheard reservation due to collision of reservation control packets during the

reservation phase is considered invalid. While resolving the reservation conflicts, the protocol should keep the control overhead at low level.
The second open issue with reservation protocols is the suitability of these protocols for a particular traffic use. Most of these protocols are generic solutions as they don't take into account the specific characteristics of multimedia traffic types such as voice or video. The issue here is related to how to set the parameters of the protocol for the given network conditions to satisfy the required QoS of a particular type of traffic.

The proposed protocol is a generic reservation scheme that can be used to provide QoS for a variety of real-time traffics. Among these traffics, we have voice. In the second part of this thesis, we apply the proposed reservation protocol to voice traffic. We propose ARPV (*Adaptive Reservation Protocol for voice traffic support over MANET*) protocol, which is an extension of our generic reservation protocol in order to take into account the characteristics of voice traffic, while enabling data traffic. Thus, we determine what are the appropriate super-frame length, the slot length, and contention parameters during the contention phase, that give the best performance for voice traffic. Among the special features of voice traffics, we have the ON/OFF characteristic due to the VAD (Voice Activity Detection). Indeed, a voice traffic source generates voice packets only during the activity (i.e., ON) period. When a voice traffic source does not have traffic to send in its assigned slot, such a slot is not used and consequently bandwidth is wasted. In order to reduce this waste of bandwidth, we propose multiplexing mechanisms where data traffic sources are enabled to share bandwidth with voice traffic sources so that bandwidth is used efficiently.

The third contribution of this thesis focuses on end-to-end bandwidth reservation and handling performance degradation due to mobility of nodes. Despite an efficient point-to-point reservation MAC scheme is a primary requirement for QoS provisioning, reserving resources along a path is also of great importance. Most reservation protocols proposed in the literature focus on point-to-point reservations, and only few works has been done to propose an efficient end-to-end reservation scheme. In fact, the task of end-to-end bandwidth reservation cannot be efficiently fulfilled without providing a tight coordination between the MAC sub-layer and the routing protocol.

We propose a reservation scheme called End-to-End Reservation scheme for Voice and data traffic support (EERV) which is an extension of our reservation MAC protocol to support the reservation and release of resources along a path in cooperation with the routing function. The particular feature of our end-to-end reservation scheme compared to existing solutions is that it does not try to reserve bandwidth only on the shortest path. Instead, it explores all other alternative paths if the shortest path does not provide the required bandwidth. This results in increasing the chance of successful end-to-end reservation at high traffic load conditions.

While setting-up reservations along a path, EERV should handle performance degradation due to mobility of nodes. When mobility of nodes is of concern, new challenges appear with reservation protocols. Thus, in a mobile environment, due to node mobility and potential change of topology, nodes may enter in the transmission range of each other resulting in collision during reserved slots. If no mechanism to handle these situations is used, collisions in one super-frame are automatically repeated in the next super-frames, resulting in frequent collisions and packet dropping. In some scenarios, mobility of nodes causes some reservation being broken without a significant change in the end-to-end path. Such scenarios can be handled by the MAC sub-layer without any action from the routing level. In other scenarios, mobility of nodes may cause reservation breakage as well as path breakage. Such scenarios need to be treated at both the MAC and routing levels. All these scenarios result in significant performance degradation, and should be carefully considered. After illustrating these scenarios, we propose mechanisms at routing and MAC levels in order to alleviate performance degradation due to mobility of nodes. These mechanisms include path breakage detection, reservation loss detection and reservation recovery.

Organization of the thesis

This thesis is organized into seven chapters. The first chapter presents the characteristics of wireless ad-hoc networks as well as an introduction to QoS provisioning over wireless ad-hoc networks. Chapter 2 presents the basic concepts of MAC protocols and a state of the art of MAC protocols in wireless ad-hoc networks.

The following chapters constitute the core of the thesis. In chapter 3, we illustrate the problem of inconsistency of reservations and reservation conflicts through different scenarios. After emphasizing the factors taking part in the occurrence of these conflicts, we propose our protocol for reservation consistency guaranteeing. The proposed protocol is a generic reservation scheme that can be used to provide QoS for different real-time traffics in wireless ad-hoc networks. In the following chapters, we extend this reservation scheme in order to take into account the characteristics of voice traffic.

Chapter 4 presents the architecture of voice communication over wireless ad-hoc networks, and the QoS requirements of voice traffic. In chapter 5, we present the ARPV (Adaptive Reservation Protocol for Voice traffic support over MANETs), which is an adaptation of our generic reservation protocol presented in chapter 4 in order to provide QoS for voice traffic. In chapter 6, we provide a performance evaluation of this protocol through a stochastic model and through simulation. Chapter 7 presents our solution for end-to-end bandwidth reservation - *End-to-End Reservation scheme for Voice and data traffic support (EERV)*. We conclude this thesis with conclusions and some perspectives.

Chapter 1: Introduction to Wireless Ad-Hoc Networks

1.1 Introduction to Wireless Communications

The field of wireless networks has grown significantly in the last three decades. Wireless networks are computer networks that use radio frequency channels as a medium of communication. Information sent by any node of the network is broadcast and can be received by all nodes in its transmission range.

The wireless communications technology has several segments such as cellular networks, Wireless LANs (WLANs), Mobile Ad-hoc Networks (MANETs), and Wireless Personal Area Networks (WPANs). Cellular networks (such as the Global System for Mobile communication (GSM)) are used to provide wide-range voice communication service. WLANs are small-scale networks (networks within a single building or campus with a size of a few kilometers) controlled by an access point, where tens (sometimes hundreds) of personal computers (PCs) are interconnected to share resources (e.g., printers), exchange e-mails, transfer files, surf the internet ... etc. A wireless ad-hoc network is a wireless network, comprised of mobile devices that share and use in a distributed way the radio channel for communication, and that do not have no fixed infrastructure (such as base station in cellular networks or access points in WLANs). Wireless ad-hoc networks can be deployed quickly anywhere and anytime as they eliminate the complexity of infrastructure setup. WPANs are the next step down from WLANs, covering smaller area with low power transmission, for networking of portable and mobile computing devices such as PCs, Personal Digital Assistants (PDAs). The particularity of WPANs is that they are designed in such a way that mobile devices consume as little power as possible so as to increase the lifetime of their batteries.

Among these networking technologies, wireless ad-hoc networks exhibit a great interest due to their potential utilization in wide range of application areas such as commercial applications, life saving and emergency systems (for example, establishing communication among rescue personnel in disaster-affected areas), and military networks. Furthermore, great moving flexibility in wireless ad-hoc networks makes such networks interesting for end-users.

In this chapter, we present fundamental concepts of wireless communications and a classification of wireless networking technologies with emphasize on wireless Ad-hoc Networks. We also provide an overview of the different approaches to provide QoS over wireless ad-hoc networks.

1.1.1 Fundamentals of Wireless Communication Technology

1.1.1.1 The electromagnetic Spectrum

The information transported on wireless networks is transmitted in the form of electromagnetic waves. These waves are characterized by their frequency and their wavelength. Table 1-1 shows various frequency bands in the electromagnetic spectrum as defined by the International Telecommunication Union (ITU).

Table 1-1. Frequency bands with their common applications.

Band Name	Frequency	Wavelength	Application
Extremely Low Frequency (ELF)	30 to 300 Hz	10.000 to 1.000 Km	Powerline frequencies
Voice Frequency (VF)	300 to 3.000 Hz	1.000 to 100 Km	Telephone communication
Very Low Frequency (VLF)	3 to 30 KHz	100 to 10 Km	Marine communication
Low Frequency (LF)	30 to 300 KHz	10 to 1 Km	Marine communication
Medium Frequency (MF)	300 to 3.000 KHz	1.000 to 100 m	AM broadcasting
High Frequency (HF)	3 to 30 MHz	100 to 10 m	Long-distance aircraft/ship communication
Very High Frequency (VHF)	30 to 300 MHz	10 to 1 m	FM broadcasting
Ultra High Frequency (UHF)	300 to 3.000 MHz	100 to 10 cm	Cellular phone
Super High Frequency (SHF)	3 to 30 GHz	10 to 1 m	FM broadcasting
Infrared	300 GHz to 400 THz	1mm to 770 nm	Consumer electronics

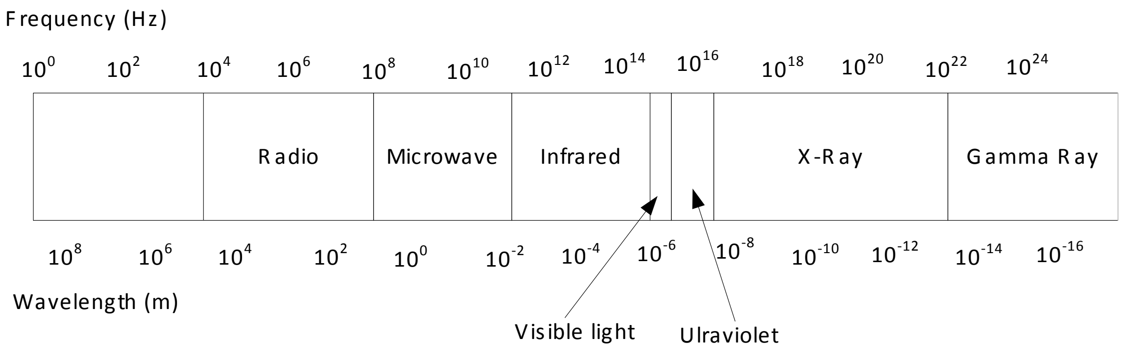

Figure 1-1. The Electromagnetic spectrum.

The low-frequency bands comprised of the radio, microwave, infrared, and visible light portion of the spectrum can be used for information transmission by modulating the amplitude, frequency, or the phase of waves. The high frequency waves such as X-rays and Gamma rays are not used due to practical concerns such as the difficulty of generation and modulation, and the harm they could cause to living beings. Furthermore, such waves do not propagate well through buildings [MS 04].

The radio waves have many interesting characteristics such as the ease of generation, their ability to pass through buildings, and the ability to travel long distances. Thanks to these characteristics, radio waves are interesting as a wireless communication medium. Thus, radio waves are widely used for both indoor and outdoor communications.

The frequency of the radio waves used for transmission has an impact on the characteristics of the transmission. Low frequency waves pass through obstacles easily, but their power falls with an inverse-square relation with respect to the distance. High frequency waves are more prone to absorption by rain drops, and they get reflected by obstacles. Because of the broadcast nature of radio waves and their long transmission range, interference between transmissions is a serious problem that needs to be addressed.

1.1.1.2 Radio Propagation

While traveling along the air, radio waves experience the following propagation phenomena [MS 04]:

- *Reflection:* When radio waves are exposed to an obstacle which is very large compared to their wavelength (such as surface of the earth, or tall buildings), theses waves get reflected by that obstacle. Reflection causes a phase shift of 180 degree between the incident and the reflected rays.

- *Diffraction:* Diffraction happens when radio waves hit an impenetrable object. The wave bends at the edges of the object, thereby propagating in different directions. The bending causes the wave to reach places behind the object which generally cannot be reached by the line-of-sight transmission.

- *Scattering:* Scattering occurs when the radio wave travels through a medium which contains many small objects when compared to the wavelength. The wave gets scattered into several weaker outgoing signals. Examples of objects that may cause scattering are street signs, lamp posts, and foliage.

1.1.1.3 Characteristics of the Wireless Channel

The wireless channel is exposed to several transmission impediments such path loss, interference, and blockage. These factors affect the transmission range, transmission rate, and the reliability of transmission. At what extend these factors affect the transmission depends on the environmental conditions and the mobility of the two communicating stations. In this section, we briefly present these factors and their impact on the reliability of transmission.

- *Path loss:* Path loss can be expressed as the ratio of the power of the transmitted signal to the power of the same signal received by the receiver. Estimating the path loss is very important in the design and deployments of wireless communication networks. Path loss depends on a number of factors such as the propagation distance, the radio frequency used, and the nature of the terrain. As these factors, and specially the terrain nature, are not the same in all environments, it is difficult to describe the characteristics of all transmissions using a single radio propagation model. Therefore, several propagation and path loss models are used to describe the different propagation environments. Among these models we have the *free space* propagation model, *two-ray* or *two-path* model, and *Nakagami* model.

- *Fading:* Fading refers to the fluctuations in signal strength when received at the receiver. Fading can be classified into two types: *fast/small-scale fading*, and *slow/large-scale fading*. Fast fading refers to the rapid fluctuation in the amplitude, phase, or multipath delays of the received signal, due to the interference of multiple versions (copies) of the same transmitted signal arriving at the receiver at slightly different times. The time between the reception of the first version of the signal and the last echoed signal is called *delay spread* [MS 04]. The main cause of fast fading are the three propagation phenomena described above, i.e., reflection, diffraction, and scattering. Slow fading occurs when objects that are located between the transmitter and receiver absorb partially the transmitted signal.

- *Interference:* When traveling on the radio channel, waves are exposed to a wide variety of sources of interference. Two main sources of interference are adjacent channel interference and co-channel interference. Adjacent channel interference refers to the interference of the on-going transmission with signals in nearby frequencies that have components outside their allocated ranges. Co-channel interference also called narrow-band interference, refers to interference with nearby systems using the same transmission frequency.

- *Doppler Shift:* The Doppler shift is defined as the change/shift in the frequency of the received signal when the transmitter and the receiver are mobile with respect to each other. If the transmitter and the receiver are moving toward each other, then the frequency of the received signal will be higher than that of the transmitted signal. Otherwise, if they are moving away from each other, the frequency of the signal at the receiver will be lower than that at the transmitter. The Doppler shift f_d is given by:

$$f_d = v/\lambda \tag{1-1}$$

  Where v is the relative velocity between the transmitter and receiver, and λ is the wavelength of the signal.

1.1.1.4 Multiple Access techniques

Due to the broadcast nature of the wireless channel, a node cannot transmit on the radio channel whenever it wants to. Multiple access techniques are used to control access to the shared channel. These techniques are based on orthogonalization of signals, where each signal is represented as function of time, frequency, or code. Hence, medium access can be performed with respect to one or multiple of these parameters, and the respective access techniques are termed frequency division multiple access, time division multiple access, and code division multiple access.

- *Frequency Division Multiple Access (FDMA):* In FDMA the total bandwidth is divided into several frequency channels or sub-bands. A transmitter-receiver pair uses a single dedicated frequency sub-band for communication. Frequency bands are separated from each other by guard frequency bands in order to eliminate the inter-channel interference. FDMA has been widely used in cellular networks. In such networks, a base station (BS) dynamically allocates frequency bands to mobile stations (MS). Each MS is allocated a pair of frequencies for communication, one for uplink communication (traffic from MS to BS), and the other for downlink communication (from BS to the MS).

- *Time Division Multiple Access (TDMA):* In TDMA, each frequency band is divided into several time slots. A set of periodically repeating time slots forms a TDMA frame. Each node is assigned one or more time slots in each frame, and the node transmits only on its dedicated slot. In two-way communications, each node is assigned two sets of slots one for uplink and the other for downlink. If the downlink and uplink slots are allocated on the same frequency band, the access scheme is called *time division duplex TDMA (TDD-TDMA)*. If they are allocated on different frequency bands, the access scheme is referred to as *frequency division duplex TDMA (FDD-TDMA)*. TDMA is widely used in second generation cellular systems such GSM.

- *Code Division Multiple Access (CDMA):* In CDMA, the narrowband signal is multiplied by a large bandwidth signal called the spreading signal. All users in a CDMA system use the same carrier frequency and may transmit simultaneously. Either TDD or FDD may be used. Each user has its own codeword that is orthogonal to all other codewords. To detect the codeword of a specific user, the receiver needs to know the codeword used by the transmitter. The receiver performs a time correlation operation of the signal with the codeword of the transmitter. Since the codewords are pairwise orthogonal, if it is the same codeword then the correlation is exact, otherwise it is zero (or, in the case of approximately orthogonal codewords, the correlation is high if it is the same codeword and low otherwise). CDMA is used in both 2G and 3G networks.

- *Space Division Multiple Access (SDMA):* SDMA provides channel access to mobile nodes based on their spatial locations. It divides the geographical space, where the users are located, into smaller spaces. Multiplexing in SDMA is performed in space through the use of directional antennas. Thus, unlike communication with omni-directional antennas where a transmission covers the entire circular region around the transmitter, a transmission with directional antenna occupies only an angular region around the transmitter. As a result, different nodes/regions that could interfere when using omni-directional communication, can communicate simultaneously when using SDMA. Thus, SDMA provides efficient use of the available bandwidth. SDMA is compatible with any multiple access scheme such as TDMA, FDMA, and CDMA.

1.1.2 Classification of Wireless Networks

Wireless Networks can be classified following different criteria: the coverage area, network topology, and operating mode.

1.1.2.1 Classification based on coverage area

A classification based on the coverage area considers the size of the covered geographical area as a main criterion. We distinguish four classes:

1.1.2.1.1 Wireless Personal Area Networks (WPAN)

The need for personal devices (such as laptops, PDAs, or smart mobile phones) to set-up short-range wireless communications with one another, without an established infrastructure, has led to the emergence of Wireless Personal Area Networks (WPANs) technology. Typically, a WPAN covers few tens of meters around a user's location, and provides the capacity to communicate and synchronize a wireless device to other wireless equipments, peripherals, and a range of pocket hardware (e.g., dedicated media devices such as digital cameras and MP3 players) [BOU 06].

Devices/users in a WPAN are totally self-organizing, that is, a device does not need any special intermediate device in order to communicate with other devices in the network. Users form temporarily network to exchange useful information, leading to a concept known as plugging in. Indeed, when any two WPAN devices are close enough (within radio communication of each other), they can communicate directly.

WPAN is a generic term referring to different technologies providing personal area networking. Examples of WPAN technologies include Bluetooth [BLUE] (IEEE 802.15.1 standard), ZigBee [ZIGBEE] (IEEE 802.15.4 standard), Ultra WideBand, and HomeRF networks [HOMERF].

1.1.2.1.2 Wireless Local Area Networks (WLAN)

Wired local area networks (LANs) have been very successful in the last few years, and now with the help of wireless connectivity technologies, Wireless LANs (WLANs) have started emerging as much more powerful and flexible alternatives to the wired LANs.

A typical WLAN contains a special node called *Access Point (AP)*, and wireless devices such as laptops and PDAs. The AP is a special node in the sense that it can interact with the other wireless devices of the WLAN as well as with an existing wired LAN such as an Ethernet. The other wireless nodes, also known as mobile stations (STAs), communicate via the AP. The AP also acts as a bridge with other networks.

1.1.2.1.3 Wireless Metropolitan Area Networks (WiMAX)

WiMAX [WIMAX] or IEEE 802.16 [802_16] is a standard for air interface for fixed broadband wireless access systems. The goal of this technology is to provide wireless access in a metropolitan area network, that is, it works as a wireless last-mile broadband access in a MAN (Metropolitan Area Network).

The WiMAX offers data rates of up to 75 Mbps per cell with each cell has a size from 2 to 10 Km. For instance, it allows the support of more than 60 channels and hundreds of DSL-type connections using a single base station [BOU 06].

1.1.2.1.4 Wireless Wide Area Networks (WWAN)

Wireless WANs are high data rate wireless networks that span a large geographical area. This category of networks includes the 2.5G (GPRS), 3G (UMTS), and 4G technologies. Thus, this technology is the most popular and most used wireless communication technology as most mobile phones are connected to a cellular network.

A classification of wireless networking technologies based on coverage area is provided in Table 1-2.

Table 1-2. Wireless communication technologies.

Category	**WPAN**	**WLAN**	**WMAN**	**WWAN**
Standard	IEEE 802.15	IEEE 802.11	IEEE 802.16	IEEE 802.20
Technologies	Bluetooth HomeRF ZigBee IR (Infrared)	WiFi HyperLAN2	WiMAX	GSM GPRS UMTS
Coverage	Tens of meters	Hundreds of meters	Tens of kilometers	Hundreds Km
Data rate	< 1Mbps	2 to 54 Mbps	Up to 70 Mbps	10 to 384 Kbps
Application	Point-to-point personal communication	Enterprise networks	broadband wireless access	Cellular and phone communication

1.1.2.2 Classification based on the infrastructure

Wireless Networks can be classified into two categories following whether an infrastructure is used or not: infrastructure-based networks, and networks without infrastructure. The GSM is an example of infrastructure-based networks. The IEEE 802.11 standard defines both modes.

1.1.2.2.1 Infrastructure-based networks

In infrastructure-based networks, all transmissions should be done through a central node called the Access Point (AP), and this even if the two communicating mobile nodes are close to each other. Generally, the AP plays the role of a router within the wireless network, or gateway that connects the wireless network with existing wired network. IEEE 802.11 WLANs [IEE 05] are typical example of infrastructure-based networks. In IEEE 802.11 WLAN, the set of stations (called STA or MT for mobile terminals) that can remain in contact (i.e., are associated) with a given AP is called a basic service set (BSS). The coverage area of an AP within which member stations (STAs) may remain connected to the AP is called basic service area (BSA). In order to be part of a BSS, mobile stations should be located within the BSA of the corresponding AP.

A BSS is considered as a basic building block of the network. Several BSSs form an extended network. Indeed, APs of these BSSs are interconnected through a distributed system (DS). The IEEE 802.11 does not specify the implementation of the DS. Thus, the DS can be of any existing network technology.

Special interconnection points called "Portals" are used in order to integrate the wireless network with existing wired networks. The BSSs, DS, and the portals together with the stations they connect constitute the extended service set (ESS) [MS 04] (cf. Figure 1-2).

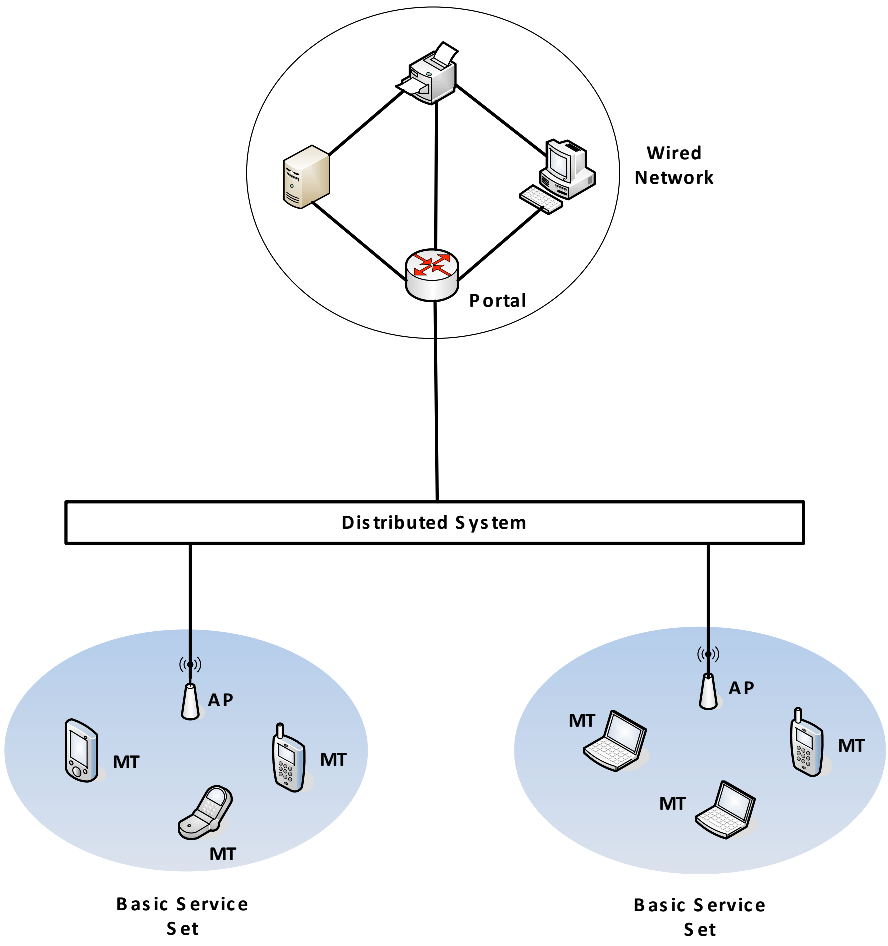

Figure 1-2. Extended Service Set.

The current infrastructure-based wireless technology extends the reach of the Internet but provides a limited mobility to users because of the limited coverage, that is, users are still restricted to stay near the base stations or access points.

1.1.2.2.2 Infrastructure less networks

Infrastructure less networks are peer-to-peer networks formed by a set of stations within the range of each other that dynamically configure themselves to set up a temporary network. Thus, unlike infrastructure-based networks, in infrastructureless configuration, no fixed controller (such as AP) is required, but a controller is dynamically elected among all the stations participating to the communication. The underlying infrastructureless property makes such configuration different from traditional WLANs, and also provides a flexible method for establishing communications in situations where geographical or terrestrial constraints demand a totally distributed network system.

In the 802.11, networks are generally implemented as infrastructure-based networks. However, the 802.11 standard also enables the construction of peer-to-peer WLANs. In this case, the IEEE 802.11 stations communicate directly without requiring the intervention of a centralized AP.

Wireless ad-hoc networks, which are the focus of this thesis, are typical example of infrastructureless networks.

1.2 Mobile Ad-hoc NETworks (MANETs)

A Mobile Ad-hoc Network (MANET) is composed of mobile nodes that communicate with each other using wireless links, and based on the peer-to-peer communication paradigm. A particular aspect of MANETs is the self-configuration of nodes. Thus, the network can have varying and arbitrary topology over the time. Each mobile node operates as a router, and is free to move randomly and connect to other nodes arbitrarily. Consequently, the network topology can change quickly and unpredictably since there may exist a large number of independent ad hoc connections. In fact, it is possible to have different applications running on the same network simultaneously [BOU 06].

An Ad hoc network is created, for example, when a group of people use wireless communications for some computer-based collaborative activities; this is also referred to as *spontaneous* networking.

Many other networking technologies can be considered as sub-classes of ad-hoc networks. For example, a Wireless Sensor Network (WSN) is a particular type of Ad-hoc Networks where nodes are characterized by slow mobility, and where the main task of nodes is monitoring or observing a phenomenon. In a Vehicular Ad-hoc Network (VANET) where nodes represent vehicles, the network is considered more predictable because vehicles are constrained by the road topology.

In this thesis, we consider the general form of wireless ad-hoc networks.

Wireless ad-hoc networks have numerous constraints related to their characteristics. It is of paramount importance to understand well these characteristics and overcome these constraints in order to provide real-time and multimedia networking over such Networks. Therefore, wireless ad-hoc networks represent an interesting research topic.

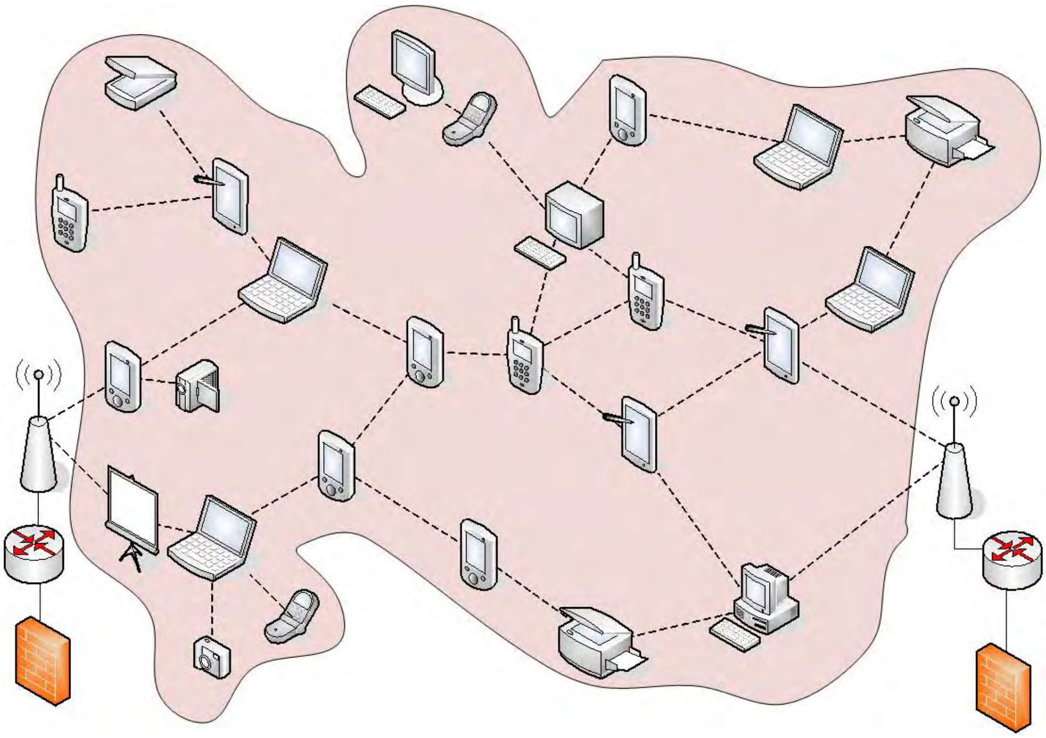

Figure 1-3. Wireless ad-hoc networks.

1.2.1 Characteristics of MANETs

Mobile ad-hoc networks have specific characteristics that make them different from the other wireless networks:

- *Wireless:* Nodes in wireless ad-hoc networks are equipped with radio interface, and use the radio channel as a communication medium. Nodes in the same neighborhood share the same medium during information exchange leading to possible collision and loss of information. Thus, the characteristics of the shared wireless medium and its consequences on the reliability of information transfer should be carefully considered.

- *Lack of infrastructure and ease of deployment:* Wireless ad-hoc networks do not need any fixed infrastructure, and consequently can be easily deployed in environments where the deployment of wired infrastructure is difficult.

- *Error-prone wireless channel:* The bit-error-rate in the wireless channel is very high (of the order of 10^{-5} to 10^{-3}) compared to that in its wired counterparts (of the order of 10^{-12} to 10^{-9}). Protocols designed for wireless ad-hoc networks should take the state of the wireless links and the signal-to-noise ratio into account.

- *Distributed and self-organizing:* Wireless ad-hoc networks do not include any centralized node that coordinates or controls the transmissions within the network. All nodes have fairly the same computing capacity, and achieve the same operations. Consequently, nodes should cooperate in order to coordinate their transmissions within the network.

- *Multi-hop networking:* Due to the limited communication range of mobile nodes, the information transmitted between two communicating nodes may cross several intermediate nodes. Thus, nodes in a wireless ad-hoc network act as routers, and ensure the relaying of information of other nodes. The set of nodes that relay the information between the source and the destination is called a *"path"*, and is determined by an ad-hoc routing protocol such as AODV [RFC 3561, PRD 00-1], DSR [RFC 4728], or DSDV [PW 94].

- *Dynamic topology:* Nodes in MANETs are free to move arbitrarily, and consequently, the network topology can change randomly and rapidly. Mobile nodes may join or leave the network at any time. In some applications, nodes may move randomly (e.g. students on a campus). In other applications such as tactical networks or in vehicular networks, mobility of nodes is more predictable.

- *Size of the network:* The number of nodes in the network can vary between few tens of nodes and few hundreds of nodes. Thus, the network size increases or decreases as nodes join or leave the network. The variable network size represents a serious problem to protocols design for Ad-hoc Networks.

- *Density of deployment:* Density of nodes can be expressed as the number of nodes per kilometer square. The density of deployment is related to the network size, and varies with the domain of application. For example, military applications require high availability of the network, making redundancy a high priority.

- *Limited resources:* Nodes in MANETs have limited resources such as computing and storage capacity. These limited resources should be carefully considered in the design of any protocol for MANETs in order to ensure their efficient utilization.

- *Limited bandwidth:* Compared to wired networks, the throughput offered by wireless networks is still considered relatively low. This limited throughput is mainly due to the limited bandwidth of the wireless channel, and represents a serious issue for multimedia applications that consume an important amount of bandwidth.

- *Limited battery power:* Some of the nodes rely on batteries or other exhaustible means of power. For these energy constrained nodes, an important system design criteria may be the energy consumption. Thus, MANET protocols should provide power control mechanisms in order to efficiently manage energy consumption of these nodes, and extend their life span.

1.2.2 Application fields

Thanks to their ease deployment and their independence of any infrastructure, mobile ad-hoc networks find their applicability in several areas. Thus, wireless ad-hoc networks tend to play an important role in new distributed applications such as distributed collaborative computing, distributed sensing applications, next generation wireless systems, and response to incidents without a communication infrastructure. Some scenarios where an ad hoc network could be used are business associates sharing information during a meeting, military personnel relaying tactical and other types of information in a battlefield, and emergency disaster relief personnel coordinating efforts after a natural disaster such as a hurricane, earthquake, or flooding.

1.2.2.1 Military and tactical networks

Among the real-life application fields in which wireless ad-hoc networks are useful, we have military tactical networks. In tactical operations and battlefield scenarios, the army is composed of groups of soldiers that need to communicate between them, and exchange useful data such as maps and terrain pictures. Deploying a fixed infrastructure for communication in such scenarios may not be feasible. Furthermore, the cellular network in the enemy territories may be controlled by the enemy and, by consequence, would not be accessible. Thus, wireless ad-hoc networks can be considered as an alternative. The applicability of wireless ad-hoc networks in the military domain has been extended to include the coordination of military objects moving at high speeds such as fleets of aircrafts or warships.

1.2.2.2 Emergency operations

Thanks to the self-configuration of the system with minimal overhead, independence of fixed or centralized infrastructure, and the random and flexibility of mobility, wireless ad-hoc networks are very useful in emergency operation. Such operations include search and rescue, fire fighting, and commando operations.

In environments where the conventional infrastructure-based communication facilities are destroyed due to a war or due to natural catastrophes such as earthquakes, the deployment of a wireless ad-hoc network can be a good solution for coordinating rescue activities and exchange coordination information.

1.2.2.3 Wireless Mesh Networks

Wireless Mesh Networks (WMNs), a special case of wireless ad-hoc networks, has been receiving much attention during the last few years. They are increasingly deployed to interoperate existing heterogeneous wireless networks (operating under infrastructure or ad-hoc modes), wired networks, and Internet, to enable mobile users with ubiquitous on-line capability and extended services provided by different networks.

A WMN is basically composed of two kinds of nodes: stationary mesh routers and mesh clients, with each node capable of operating as both a host and a router. Mesh routers have minimal mobility and form the mesh backbone of the WMN. They are self-configurating and self-healing to robustly interoperate various existing networks such as Internet, cellular, Wi-Fi, WiMax, and sensor networks. Thus, they act as gateways/repeaters, and their main role is to relay traffic from mesh clients and other routers to the internet backbone. Mesh clients are the end points of a mesh network and can be either stationary or mobile. They may form a client mesh network attached to fixed mesh routers, and rely on mesh routers to access the internet services. Some mesh routers are connected to the internet by wired connectivity, while the rest of nodes access the internet through these wired-connected routers forming multi-hop network with them.

As example, a wireless mesh infrastructure can be formed by placing small radio relaying devices on the rooftops of the houses in a residual zone. These radio equipments self-organize their communications to form a multi-hop network. The residents can connect to any one of these equipments in order to get low cost internet connection. Indeed, the cost of deploying a Wireless Mesh Network is much less than the one required for cellular network counterparts [BOU 06].

1.2.2.4 Wireless Sensor Networks

Wireless Sensor Networks (WSN) are a special category of ad-hoc networks that are used to provide a wireless communication infrastructure among nodes, called sensors, deployed in a specific application domain. Sensors are small devices equipped with wireless interfaces that have the capability of sensing physical parameters, data processing, and communicating with other sensors. Indeed, the main task of sensors is to monitor a specific phenomenon, and deliver useful information about the monitored area to a special node called *"sink"*. An example of such tasks is the measurement of parameters such as temperature, humidity, and nuclear radiation. Examples of domains of application of WSN include cold chain monitoring, leak detection, home security, health care, and environmental monitoring.

WSN have special features that make of them a distinct category of ad-hoc networks. These features are mainly:

- *Mobility of nodes:* Mobility is not a mandatory requirement in sensor networks. The mobility pattern of sensor nodes depends on the application. For example, sensors deployed in a vegetables stock for periodic monitoring of the temperature, are not required to be mobile. Contrarily, sensors deployed for cold chain monitoring may be designed to support mobility.

- *Size of the network and density of deployment:* The number of nodes in a sensor network and their density can be much larger than in a typical ad-hoc network. Thus, as sensors have small size, it is possible to deploy hundreds even thousands of nodes on a small area easily.

- *Power constraint:* Sensor nodes are expected to operate in hard environmental conditions, where the human action is minimal if not impossible. In some environments, recharging batteries or their substitution is impractical.

1.3 QoS in Wireless Ad-hoc Networks

Wireless ad-hoc networks present many interesting features such as moving flexibility, low cost, and easy deployment. With their deployment, all the services existing on the internet such as Voice over IP (VoIP) and multimedia applications should become available for the MANETs users. Thus, in addition to data centric applications such as web browsing, file transfer, and e-mail exchange, wireless ad-hoc networks should have the ability to provide cost effective multimedia service, allowing the integration of multimedia and data communication over the same network.

The types of multimedia services that run over a wireless ad-hoc network depend on the application field. For instance, in emergency operations such as firefighting and life-saving, voice communication is vital in order to coordinate rescue activities. Tactical navigation requires quick, secure, and reliable multimedia multicasting. For example, the leader of a group of soldiers may need to give an order to all or a sub-set of its group members. Also, the group members may need to exchange videos about the ground in real-time. Hence, in this application, the network should be able to provide secure and reliable multicast communication, with the support of real-time traffic.

Multimedia applications require some performance level of a service offered by the network that is called *Quality of Service (QoS)*. The aim of QoS provisioning is to achieve a more deterministic network behavior, so that the data carried by the network can be better delivered and network resources can be better utilized [SM 04]. A network can provide different levels of services to the users. Thus, a service can be characterized by a set of measurable service parameters such as the minimum guaranteed bandwidth, maximum tolerable delay, maximum delay variance (or jitter), and the maximum packet loss rate. When the network accepts a service request from the user for a traffic flow, it should guarantee that the service requirements of the user are met.

In fact, QoS provisioning is a challenging problem in wireless ad-hoc networks as well as in other networking technologies such as wired and cellular networks. However, the support of multimedia applications over wireless ad-hoc networks is more challenging and requires more complex solutions due to the special characteristics of such networks (cf. section 1.2.1), and since the bandwidth in these networks is much limited than that in wired counterparts.

Indeed, the physical, MAC, and routing layers designed for wireless ad-hoc networks are still unable to provide efficient utilization of the limited bandwidth. For example, the IEEE 802.11 [IEE 05], which is a de-facto standard, cannot fulfill the QoS requirements of multimedia since it was not originally designed to support delay-sensitive applications. Thus, QoS provisioning for multimedia traffic support over wireless ad-hoc networks is still a challenge, and an efficient solution for multimedia/data integration in such networks is still an open research field.

To provide an effective solution for multimedia/data communication over wireless ad-hoc networks, it is necessary to understand the characteristics of multimedia traffics, their QoS requirements, and the main issues in QoS provisioning for these traffics in wireless Ad-hoc networks.

1.3.1 Issues and challenges in QoS provisioning over MANETs

Wireless ad-hoc networks have unique characteristics that represent several issues in providing Qos. Some of the characteristics that affect QoS in such networks are the dynamic varying network topology, lack of control controller, error-prone shared wireless channel, limited resource availability, insecure medium, and lack of precise state information. Below, we briefly discuss the impact of each of the above-mentioned characteristics on QoS provisioning in wireless ad-hoc networks.

- *Dynamically changing network topology:* An ad-hoc network topology changes dynamically because nodes have no restriction on mobility. Thus, admitted QoS sessions may suffer due to frequent path breakages, thereby requiring such sessions to be reestablished over new paths. The delay required to reestablish a broken session may cause some of the packets belonging to the session to miss their deadline, which is not acceptable for applications with stringent delay requirements.

- *Lack of central coordination:* Unlike wireless LANs and cellular networks, wireless ad-hoc networks do not have central controller that coordinates the communication between nodes forming the network. This complicates further QoS provisioning in wireless ad-hoc networks.

- *Error-prone shared wireless channel:* The wireless channel that represents the primary medium of transmission for MANETs is exposed to several impairments such as reflection, diffraction, scattering and interference with other radio technologies. These imperfections lead to a high bit-error-rate, and consequently to high packet loss rate, which is not acceptable for multimedia applications.

- *Limited resource availability:* Resources such as bandwidth, battery, storage capacity, and processing capability are limited in wireless ad-hoc networks. Out of these, bandwidth and battery life are critical resources, the availability of which significantly affects the performance of the QoS provisioning mechanism. Hence, efficient resource management mechanisms are required in order to ensure the optimal utilization of these resources.

- *Insecure medium:* Communication through the radio channel is highly insecure due to the broadcast nature of the wireless medium. Therefore, security is an important issue in wireless ad-hoc networks. Since they use the radio channel as transmission support, wireless ad-hoc networks are subject to attacks such as eavesdropping, spoofing, denial of service, message distortion, and impersonation. Without sophisticated security mechanisms, it will be very difficult to guarantee secure communications.

- *Imprecise state information:* Nodes in wireless ad-hoc networks maintain some state information about the network such as link-specific state information or flow-specific state information. Link state information includes bandwidth, delay, delay jitter, loss rate, error rate, and stability of the link. The flow state information includes the flow identifier, source/destination addresses, and QoS requirements (such as maximum and minimum required bandwidth, maximum delay, and maximum delay jitter). This state information is imprecise due to the dynamic network topology and channel characteristics.

These issues and others make of QoS provisioning in wireless ad-hoc network a challenging issue. Hence, QoS provisioning requires several schemes such as service negotiation, resource reservation, priority scheduling, and call admission control. In the next section, we give a brief classification of the different QoS provisioning approaches in wireless ad-hoc networks.

1.3.2 Classification of solutions for QoS provisioning over MANETs

In the literature, the research on QoS support in MANETs spans over all the layers in the network. Each layer has been the focus of many research works. In this section, a brief discussion of QoS provisioning approaches in wireless ad-hoc networks is provided.

1.3.2.1 QoS models

A QoS model specifies an architecture in which some kinds of services could be provided. Several service models have been proposed. Three of these models are the Integrated Service (*IntServ)* model [RFC 1633], the differentiated service (*DiffServ*) [RFC 2475], and the Flexible QoS Model for mobile ad-hoc networks (FQMM) [XIA 00].

The *IntServ* model provides QoS on a per flow basis, where each node maintains a state for each flow, specific state information such as bandwidth requirements, delay bound, and cost. The model defines three types of services, namely, guaranteed service, controlled load service, and best-effort service. The resource reservation protocol (*RSVP*) [RFC 2205] is used for reserving the required resources along a path. Due to the high volume of information maintained by IntServ-enabled nodes, the model can be applied only for small-sized ad-hoc networks, and may not scale for large networks. In addition, due to the limited resources of mobile nodes and the frequent changes of the network topology, it would be difficult to maintain a per-flow information in each mobile node.

The *DiffServ* model was designed in order to overcome the difficulty in implementing and deploying IntServ and RSVP. While IntServ provides per-flow guarantees, *DiffServ* follows the philosophy of mapping multiple flows into a few service classes. At the boundary of the network, traffic entering a network is classified, conditioned and assigned to different behavior aggregates by marking a special *DS (Differentiated Services)* field in the IP packet header (TOS field in IPv4 or CLASS field in IPv6). Within the core of the network, packets are forwarded according to the per-hop behavior *(PHB)* associated with the DSCP *(Differentiated Service Code Point)*. This eliminates the need to keep any flow state information elsewhere in the network.

The *IntServ* and *DiffServ* models cannot be directly applied to wireless ad-hoc networks because of the inherent characteristics of such networks such as varying network topology, limited resource availability, and error-prone wireless channel.

Flexible QoS Model for Mobile ad-hoc networks (FQMM) [XIA 00] is a hybrid service model that takes into account the characteristics of Wireless Ad-hoc networks. The basic idea of that model is that it uses both per-flow state property of *IntServ* and the service differentiation of *DiffServ*. In other words, this model proposes that highest priority is assigned per-flow provisioning and other priority classes are given per-class provisioning. It is based on the assumption that not all packets in the network seek for highest priority. The FQMM model defines three types of nodes, exactly as in *DiffServ* a) ingress, b) core and c) egress. The difference though is that in FQMM the type of a node has nothing to do with its physical location in the network, since this wouldn't make any sense in a dynamic network topology. A node is characterized as ingress if it is transmitting data, core if it is forwarding data and egress if it is receiving data.

1.3.2.2 QoS resource reservation signaling

The QoS resource reservation signaling scheme is responsible for reserving the required resources and informing the corresponding applications, which then initiates data transmission. Signaling protocol consists of three phases: connection establishment, connection maintenance, and connection tear down. The signaling scheme ensures that an established connection meets its QoS requirements for all the duration of the connection, and repairs/reconfigures the path if the connection suffers from any violation of its QoS guarantees. At the end of the session, the signaling scheme releases resources that had been reserved for that connection.

In wired networks, RSVP protocol [RFC 2205] is used as resource reservation signaling scheme. However, RSVP cannot be directly applied to wireless ad-hoc networks because resources are assumed to be available nearly for the application throughout the session once resources are reserved, the think which is not true in wireless ad-hoc networks due to the instability of the network. Furthermore, the control overhead generated during the connection maintenance phase of RSVP is too heavy for ad-hoc networks [MS 04].

MRSVP protocol [TBA 01], is an extension to RSVP for mobile networks. MRSVP assumes that mobile host can connect to the network through different access points during its connection. The host is assumed able to predict the set of locations and access points that it is expected to visit during the connection. Based on these assumptions, MRSVP establishes two types of reservations: *active reservations* and *passive reservations.* One active reservation is established between the current path, i.e., between the access point to which the mobile host is currently attached and its corresponding destination. Many Passive reservations are established on the other alternative paths, i.e., paths toward access points that the mobile host is expected to visit in the future. Resources that are reserved passively for a traffic flow can be used by other flows that require best-effort service.

MRSVP is an interesting solution for resource reservation in mobile environments. However, its applicability in wireless ad-hoc networks is challenging because it requires the prediction of the future locations of mobile hosts in advance, the thing which is not obvious in wireless ad-hoc networks.

1.3.2.3 QoS routing protocols

Routing is an essential component to realize a complete QoS wireless ad-hoc network Architecture. Routing protocol is responsible of finding a path to be followed by data packets from a source node to a destination node.

Best-effort ad-hoc routing protocols try to find paths and react to mobility of nodes without any consideration of bandwidth availability on these paths. Thus, the main criterion in route calculation is the path length where shortest paths are privileged. A variety of best-effort routing protocols for wireless ad-hoc networks have been proposed in recent years, such as Ad hoc On-demand Distance Vector routing (AODV) [RFC 3561], Dynamic Source Routing protocol (DSR) [RFC 4728, JMB 96, JMB 01], Optimized Link State Routing protocol (OLSR) [RFC 3626], and Destination Sequenced Distance Vector (DSDV) routing protocol [PW 94]. These algorithms differ in the way routes are found and the way they react to network topology changes.

QoS routing protocols in wireless ad-hoc networks extend best-effort routing protocols, and search for routes with sufficient resources in order to satisfy the QoS requirements of traffic flows. The information regarding the availability of resources is managed by a resource management module which assists the QoS routing protocol in its research of paths that satisfy the required bandwidth. Ticket-based QoS routing [CN 99], Predictive Location-based QoS Routing protocol [SN 02], Trigger-based Distributed QoS Routing (TDR) protocol [DE 02], QoS-enabled Ad-hoc On-demand Distance Vector routing protocol (QoS AODV) [PRD 00-2], and On-Demand Link-State Multipath QoS Routing protocol [CHE 02] are examples of QoS routing protocols.

Despite the variety of research in the field of QoS routing in wireless ad-hoc networks, end-to-end QoS provisioning is still a challenging task. The addition of bandwidth availability metric in route selection complicates further the task of QoS routing because bandwidth availability information fluctuates as the network topology changes. Thus, a path currently satisfying the required bandwidth may not guarantee the same available bandwidth in the near future. In addition, as path breaks occur frequently in wireless ad-hoc networks, the path satisfying the required QoS need to be recomputed every time the current path gets broken. QoS routing protocols should respond quickly to path breaks and recompute the broken path or bypass the broken link without degrading the level of QoS, the task which is not easy to achieve in wireless ad-hoc networks.

1.3.2.4 QoS MAC protocols

The MAC protocol in wireless ad-hoc networks determines which node is allowed to use the wireless medium next when several nodes are competing for transmission. Existing MAC protocols for wireless ad-hoc networks use channel sensing and random backoff schemes, providing opportunistic medium access, thus, making them suitable for best-effort traffic.

The most widely deployed medium access technology is the IEEE 802.11 standard [IEE 05], which has two modes of operation: a distributed coordination function (DCF) mode, and a point coordination function (PCF) mode. The DCF mode provides best-effort access service, while the PCF mode is designed to provide real-time traffic support in infrastructure-based wireless networks. Due to the lack of fixed infrastructure in wireless ad-hoc networks, the PCF mode cannot be used to provide real-time service in wireless ad-hoc networks.

The IEEE 802.11e and the EDCF scheme [IEE 05] are extensions of the legacy 802.11 standard to support QoS in wireless ad-hoc networks through a differentiated access scheme. Despite this service differentiation at the MAC level, the IEEE 802.11e does not ensure the QoS for real-time traffic. Hence, supporting real-time traffic in these networks requires a more deterministic access.

The legacy version of the 802.11 and its extension are discussed in detail later in chapter 2. In addition to these standardized MAC protocols, several other MAC protocols for the support of QoS for real-time and multimedia applications in wireless ad-hoc networks have been proposed. Some of these protocols are discussed in chapter 2.

1.3.2.5 QoS frameworks

A QoS framework is a complete system that attempts to provide the required service to each user or application. The key component of any QoS framework is the QoS model which determines the approach through which the user requirements are met. For example, a framework may serve users on a per flow basis (the case of *IntServ* model), or on a per class basis (the case of *DiffServ*). The other components of a framework are: QoS signaling for resource reservation, QoS routing in order to find feasible paths that satisfy the required bandwidth, QoS medium access control, call admission control, and packet scheduling schemes. All the components of the framework should cooperate together in order to satisfy the user service requirements.

Many QoS frameworks had been proposed in the literature. Examples of such frameworks include INSIGNIA [SEO 00], INORA [DHA 02], SWAN [GAH 02], and Proactive RTMAC [VIV 04].

1.4 Summary

Wireless ad-hoc networks represent many interesting features. The greatest advantage of such technology is its easy and cost effective deployment. The coverage area can be scalable by adding to or removing nodes from the ad-hoc network. The functions of self-organization, auto-configuration, and self-healing are intrinsic to wireless Ad-hoc Networks.

In this chapter, we presented the basic concepts and characteristics of wireless ad-hoc networks, with more emphasis on QoS and the principal issues in providing QoS over wireless ad-hoc networks.

Due their quick and economically less demanding deployment, wireless ad-hoc networks find applications in several areas such as wireless mesh networking, wireless sensor networks, military applications, and emergency operations.

Depending on the application field, wireless ad-hoc networks users expect certain level of service from the network such as minimum bandwidth, maximum delay, and maximum packet loss rate. After receiving a service request from the user, the wireless ad-hoc network architecture should ensure that the service request of the users is met. However, the characteristics of wireless ad-hoc networks such as dynamic network topology and instability of the radio channel make QoS provisioning a very challenging task.

The goal of QoS provisioning is to define mechanisms in order to achieve a more deterministic network behavior so that the user QoS requirements are satisfied. Research community is working since nearly a decade with the aim of proposing efficient protocols for QoS provisioning in wireless ad-hoc networks. Throughout our readings, we found that while all the layers of the network contribute in enforcing the QoS offered to the end-users, the MAC sub-layer has been always considered as the core of the problem. Thus, QoS can be provided to end users only if an efficient QoS-aware medium access scheme is provided at the sub-MAC layer.

In the following chapter, we provide a survey of MAC protocols proposed in the literature. We expose the major challenges that face the design of efficient medium access schemes, and the effort that has been made to enhance their performance in order to satisfy the QoS requirements of real-time and multimedia applications.

Chapter 2: MAC protocols for Ad Hoc Networks: Concepts, Issues, and Taxonomy

2.1 Introduction

Multimedia and real-time applications require strict QoS in terms of bandwidth and delay guarantees. Regarding their functions of controlling the share of limited wireless bandwidth, Medium Access Control (MAC) protocols play an important role to fulfill these QoS requirements in MANETs.

The design of QoS MAC for Ad hoc Wireless Networks is a challenging task for several reasons. The wireless channels are time-varying and error-prone. So, using radio as the transmission carrier, the transmitted information is strongly affected by the surrounding environment and sources of interference such as reflection, diffraction, and scattering. The variation of the environment due to nodal mobility and changes of topology, combined with sources of interference, leads to a complex and time-varying communication environment in which it is difficult to provide QoS. Furthermore, the lack of the central control entity makes QoS MAC design even harder since most existing QoS mechanisms and algorithms, such as fair queuing and scheduling, require the presence of a central control point.

In this chapter, we provide a comprehensive panorama of MAC protocols in MANETs. A classification of the major classes of MAC protocols proposed in the literature is presented. A number of representative MAC protocols for each class are highlighted, with a discussion of their major advantages and issues in regard to QoS provisioning.

2.2 Issues in designing MAC protocols for MANETs and sources of impairments

The following are the main issues that need to be addressed while designing MAC protocols for MANETs.

2.2.1 Lack of centralized coordination

Unlike in cellular networks where access to the wireless medium and resource allocation are controlled by a centralized entity (i.e., the base station), MAC protocols design for wireless ad-hoc networks is a more challenging task because of the need of distributed access, node mobility and lack of coordination between mobile nodes. Since there is no coordinator that controls the access to the shared medium, nodes are required to self coordinate their transmissions and access to the medium through the exchange of control information. However, the control traffic exchanged between mobile nodes may be high causing a significant waste of bandwidth and processing overhead. Thus, the MAC protocol design should provide a fair and distributed access with keeping the overhead at low level.

2.2.2 Half-duplex radio transceiver

Wireless ad-hoc networks use half-duplex communication to reduce cost and complexity of the transceiver. In *half-duplex* radio communication systems, the communication is two-way, but the same transceiver is used for both transmission and reception on the same channel. Consequently, at any point in time, a node can use its transceiver either to transmit or to receive, and not for both operations. Thus, when a node is using its transceiver for transmission, it cannot take into account any other operation (transmission or reception) taking place on the wireless channel. However, when a node is using its transciever for reception, any other transmission in its transmission range may result in collision of the packets currently being received.

2.2.3 Shared broadcast channel

In contrast to point-to-point channels (e.g., point-to-point links in wired networks), the radio channel is broadcast, i.e., the signal transmitted by a node is received by all the nodes in its transmission range. This can be an advantage if a packet is addressed to all neighboring nodes; a broadcast may be achieved by a single transmission (cf. Figure 2-1). However, in unicast transmissions, the broadcast channel is an obstacle against achieving high spatial reuse. The spatial reuse refers to the possibility of simultaneous transmissions on the same channel without causing interference when nodes are sufficiently far apart.

To explain this problem, consider Figure 2-2 where node *n1* is sending a packet to node *n2*. The correct reception of the packet by *n2* requires that at the moment where *n1* is sending its packet, no other node in the neighborhood of *n2* is transmitting simultaneously. Otherwise, any simultaneous transmission causes collision of the packet transmitted by *n1*. In addition, no node can transmit a packet to the neighbors of *n1* until *n1* finishes its transmission. Thus, unfortunately, node *n3* is prohibited from transmitting a packet to *n4* despite it is not engaged in any packet transmission or reception. Similarly, nodes *n6* and *n8* cannot receive any packet from *n5* and *n7* respectively despite they are not engaged in transmission or reception.

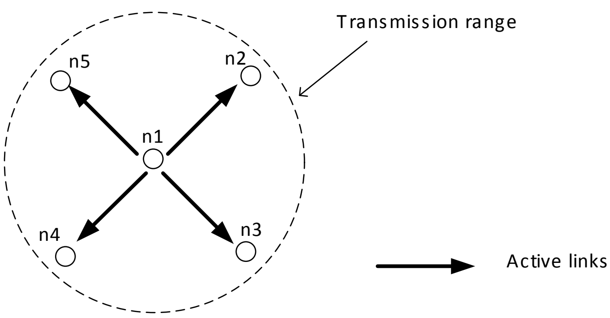

Figure 2-1. Advantage of the broadcast channel in broadcast transmissions.

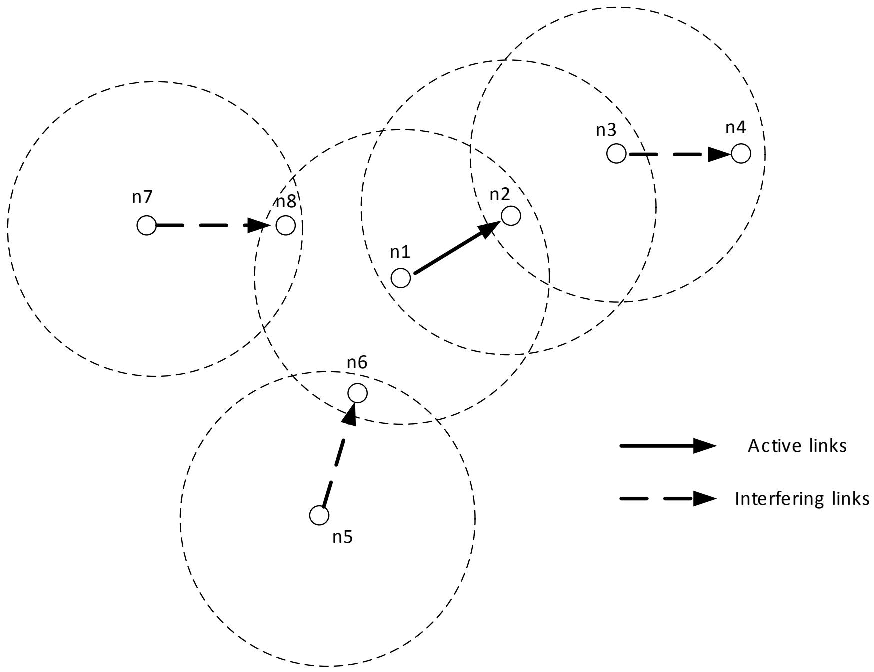

Figure 2-2. Impact of broadcast channel on the spatial reuse (for readability we do not show the transmission range of nodes *n6* and *n8*).

The spatial reuse can be increased through the use of directional antennae technology, which is out the scope of this thesis.

2.2.4 Hidden and exposed terminal problems

Hidden and exposed terminal problems are inherent to wireless ad-hoc networks. The hidden terminal problem happens when two or more nodes which are not in the transmission range of each other transmit packets to a common neighboring node simultaneously. Transmitted packets collide at the receiver. The senders, unaware of this, may get the impression that the receiver can clearly listen to them without interference from any else. Unfortunately, the senders will continue their transmission, and will be aware of the collision only when they finish their transmission without receiving an acknowledgement.

Figure 2-3 shows the hidden terminal problem. Suppose that node *n1* is currently transmitting a packet *P1* to node *n2*. Since *n3* is not in the transmission range of *n1*, it cannot hear the transmission from *n1* to *n2*. *n3* initiates the transmission of its packet *P2* to *n2*, and collision occurs at *n2*. Both *n1* and *n3* which are not aware of the collision continue sending their packets in vain.

The hidden terminal problem is due to the fact that a node has no information about the transmissions of its two-hop neighbors. At high traffic load, the probability of simultaneous transmissions is high, and consequently the probability of packet collision is quite important. Thus, the hidden terminal problem reduces significantly the network throughput, especially when the traffic load is high.

A way to resolve the hidden terminal problem consists in making a transmitting node informing its two-hop neighbors of its current transmission so that simultaneous transmissions with these neighbors is avoided. However, depending on the number of control messages required to inform these neighbors about the future transmission, the control overhead of this scheme may be high.

The exposed terminal problem is illustrated in Figure 2-4. Suppose that *n2* is transmitting a packet to *n1*. *n3* is in the transmission range of the sender, *n2*, but out of the transmission range of the receiver, i.e., *n1*. *n3* defers transmission to *n4* upon hearing a packet from *n2*. In this case *n3* is refrained from transmission even though its transmission does not interfere with the reception at *n1*. *n3* in this scenario is called exposed terminal.

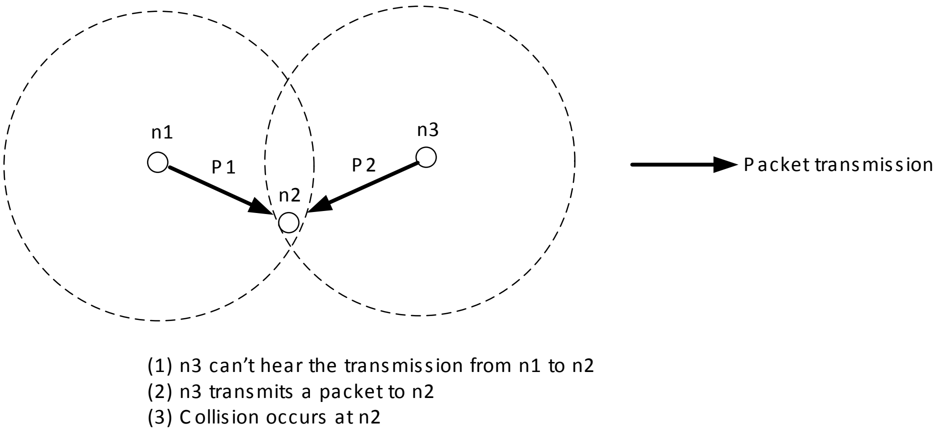

Figure 2-3. Hidden terminal problem.

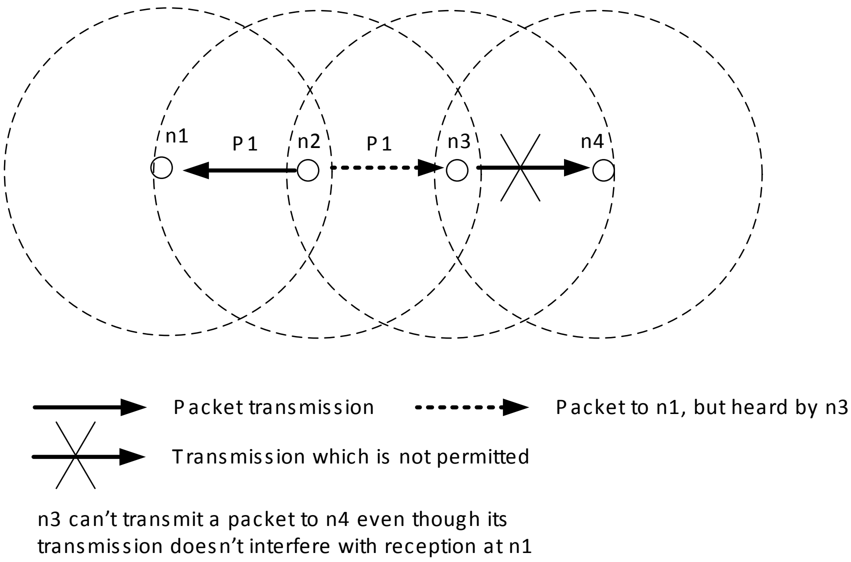

Figure 2-4. Exposed terminal problem.

The exposed terminal problem reduces the spatial reuse and results in bandwidth being underutilized. While potential solutions have been proposed for the hidden terminal problem, satisfactory solutions to the exposed terminal problem still do not exist. MAC protocols must provide solutions to resolve both the hidden and exposed terminal problems, while avoiding the increase of the control overhead.

2.2.5 Bandwidth efficiency (limited bandwidth)

Future wireless networks are expected to provide users with considerably higher data rates than those offered today. For example, the IEEE 802.11a and 802.11b wireless standards [IEE 05], offer interesting data rates of 54 and 11 Mbps respectively. Despite these data rates, the effective throughput offered by the network to traffic flows is still very low. Thus, from the 11 or 54 Mbps data rate offered by the wireless channel, only a little amount is used for traffic transmission. This is mainly due to the broadcast nature of the wireless channel, high packet collision rate, the hidden and exposed terminal problems, and bandwidth consumed by the control overhead.

Since the wireless channel (the air) is shared, only one node can use the channel for transmission in its transmission range at any time. This node monopolizes the available bandwidth offered by the wireless channel until the end of the transmission, thus reducing the possibility of spatial reuse and global network throughput.

Collisions contribute further in reducing the efficient usage of the limited available bandwidth, because the bandwidth which is used for the transmission of collided packets is wasted. Hidden and exposed terminal problems are considered harmful for efficient bandwidth utilization as they contribute in increasing the collision rate.

The control traffic overhead such as the RTS and CTS packets and acknowledgments used in the IEEE 802.11 standard, results in a significant waste of the available bandwidth.

The MAC protocol design cannot act on the broadcast nature of the wireless channel, but can achieve efficient use of the bandwidth through lowering the control traffic overhead and packet collision, and reducing the effect of the hidden and exposed terminal problems.

2.2.6 Heterogeneous Quality of service requirements

Wireless ad-hoc network infrastructure is expected to serve different kinds of applications with different QoS requirements, varying from reliable file transfer to real-time multimedia such as voice live conversations and video streaming. This heterogeneity adds more difficulty in MAC protocols design. Thus, in addition to traditional throughput, each application has its own required latency, delay jitter, and tolerable loss rate.

Thus, a MAC protocol which is designed to satisfy the QoS requirements of a particular type of traffic may not be suitable for the other traffic types. A MAC protocol that satisfies the QoS requirements of heterogeneous traffic types is thus a great challenge.

2.3 Design goals of MAC protocols for MANETs

The following are the most important goals to meet while designing a QoS-aware MAC protocol for wireless ad-hoc networks [SM 04]. These points are also used as criteria of evaluation of the performance of existing MAC protocols.

- The operations of the protocol should be distributed.
- The protocol should provide QoS support for real-time traffic.
- The access delay, which refers to the average delay experienced by any packet at the MAC sub-layer to get transmitted, must be kept low.
- The protocol must use bandwidth efficiently.
- The protocol should ensure fair allocation of the available bandwidth to nodes.
- The control overhead should be kept low.
- The protocol should reduce the effect of hidden and exposed terminal problems.

2.4 Taxonomy of MAC protocols in Wireless Ad-hoc Networks

2.4.1 Best-effort access schemes: IEEE 802.11 standard basics

Today, IEEE 802.11 [IEE 05] technology is dominant: most existing WLAN (Wireless Local Area Networks) and wireless ad-hoc networks are generally based on the IEEE 802.11 standard. Furthermore, most MAC protocols designed for QoS support are extensions or adaptations of IEEE 802.11. In this section, we recall the basic mechanisms of IEEE 802.11 standard that are useful to understand the protocols presented in the other sections to provide QoS support.

The IEEE 802.11 considers two network architectures: ad-hoc and infrastructure-based. In an ad-hoc network, mobile terminals communicate with each other without the need of any infrastructure. In infrastructure-based networks, mobile terminals communicate through an Access Point (AP). The AP manages all communications between mobile terminals belonging to the same area which is called BSS (Basic Service Set).

The IEEE 802.11 standard defines two access modes: a distributed access mode called Distributed Coordination Function (DCF) used in MANETs, and a centralized contention free mode called Point Coordination Function (PCF). The DCF mode provides best-effort service, while PCF mode is designed to provide real-time traffic support in infrastructure-based Wireless LANs. The protocol operates on a super-frame basis. The super-frame is composed of a Contention Free Period (CFP), followed by a Contention Period (CP). The DCF is the basic access scheme in the CP, while the PCF is the access scheme in the CFP.

DCF mode is based on the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). It uses two mechanisms to avoid collision: the physical carrier sensing, and the virtual carrier sensing. The physical carrier sensing is used to detect the presence of signal on the common shared physical channel. The virtual carrier sensing uses the duration field of the MAC header to indicate the duration during which a station will hold the channel.

Data transmission in DCF is accomplished following the RTS/CTS/DATA/ACK handshake illustrated in Figure 2-5. A station, which has a DATA packet to send, waits the channel to remain idle for the duration of DIFS (DCF Inter Frame Space). If the channel lasts idle for DIFS, the station transmits an RTS packet. Otherwise, the station enters in a backoff period, by choosing a backoff timer uniformly distributed in [0, CW], (CW is the Contention Window). The backoff timer is decreased by one for each idle time-slot corresponding to one-way propagation delay. The station transmits its RTS packet when the backoff timer expires. When the receiver successfully receives the RTS packet, it waits for SIFS (Short Inter-Frame Space) before replaying with a CTS packet. Both the RTS and CTS packets contain the Duration field, which is used in order to prevent neighbors from accessing the channel during the RTS/CTS/DATA/ACK handshake. When the sender receives the CTS packet, it sends the DATA packet after SIFS period. The receiver acknowledges the reception of DATA packet by sending an ACK packet after SIFS period.

In order to avoid the hidden terminal problem, each station includes a Network Allocation Vector (NAV) timer. The NAV specifies the duration for which the access to the channel is not allowed in order not to jam the pending transmissions. When the neighbors of the sender or the receiver receive the RTS, CTS or Data packets, they update their NAV with the duration specified in the received packet, and differ their access to the channel for the duration specified (in the NAV timer). This mechanism is called Virtual Carrier Sensing.

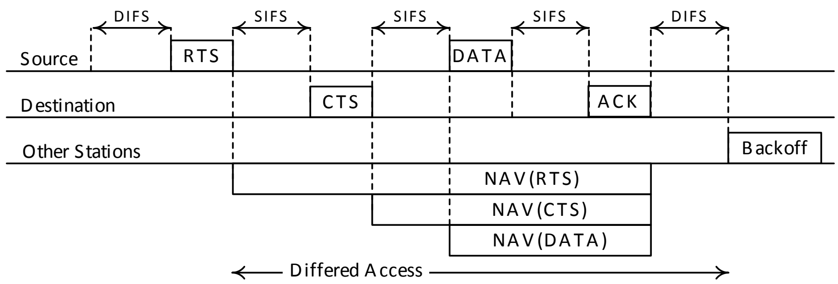

Figure 2-5. Distributed Coordination Function illustration.

The major advantage of the IEEE 802.11 standard is its simplicity. However, it does not provide any guarantees for traffics with stringent QoS requirements.

2.4.2 QoS oriented MAC protocols

QoS limitations of IEEE 802.11 standard motivated numerous research efforts to enhance its performance, and extend its functionalities in order to make it QoS-aware. Among these extensions, prioritization schemes and resource reservation schemes are the most investigated directions toward MAC level QoS provisioning. Other approaches that consider the use of directional antennae or multiple channels are out of the scope of this thesis.

2.4.2.1 Prioritization-oriented MAC protocols

Prioritization protocols control the inter-frame spacing and backoff values to meet the delay and bandwidth requirements of real-time traffic. These schemes are relatively straightforward extensions of IEEE 802.11 DCF and can be overlaid on this protocol. In this section, we review the most representative prioritization-oriented protocols.

RT-MAC protocol

In the basic IEEE 802.11 protocol, packets are transmitted (and retransmitted in the event of collision) even if their deadlines are already missed (because, the protocol does not take any care of the deadlines). This behavior leads to a waste of resources. RT-MAC (Real-Time MAC) protocol [BDM 99, BDM 01] has been proposed to reduce the number of packets collisions and the transmission of packets missing their deadlines. RT-MAC introduces two additional mechanisms to achieve its goals. Packets scheduling based on packets deadlines, and an enhanced collision avoidance scheme to determine the transmitting station's next backoff value (BV).

When a real-time packet is submitted for transmission, a deadline is associated with the packet. The deadline of a packet is examined at three key points to determine whether to discard it or not. A packet is first examined 1) when it is removed from the packet queue for transmission, 2) when the backoff timer expires, and 3) when a transmission goes unacknowledged. During one of these three points if a packet exceeds its transmission deadline, it is automatically discarded otherwise the packet is transmitted. By discarding a packet as soon as its deadline is exceeded, transmission queue length is kept low, and as a result, the chance that other packets in the queue will meet their deadlines is increased.

In order to avoid successive collisions, RT-MAC forces the stations to use different values of the backoff timer. This is realized by time-stamping transmitted packets with the next BV to be used by the transmitting station. Stations that hear the transmission use this BV to avoid selecting the same BV for their backoff timer. This ensures that neighboring stations use different backoff values for their transmissions and consequently the number of collisions is reduced. RT-MAC scheme has been shown to achieve drastic reductions in mean packet delay, missed deadlines, and packet collisions as compared to the basic IEEE 802.11.

DCF-PC protocol

The DCF-PC (DCF with priority classes) [DC 99] was proposed as an extension to the IEEE 802.11 standard in order to support different classes of traffic. The main idea of DCF-PC is that priority-based access to the wireless medium is controlled using different inter-frame space (IFS) time intervals. Prioritization is made through assigning shorter IFS and shorter random backoff time for higher priority stations. While normal stations wait for the channel to remain idle after DIFS interval before they transmit data, a higher priority station waits only for PIFS (Point Inter Frame space). PIFS, defined in the IEEE802.11 standard, specifies the inter-frame used by the AP to access the channel, and is smaller than DIFS and IFS. The priority scheme of DCF-PC is illustrated in Figure 2-6. After station A finishes its transmission, stations A, B, C and D compete to get access to the channel. Station C, which has higher priority than the other stations waits the channel to remain idle during PIFS and starts its backoff timer. The other stations wait the channel to be idle for DIFS before they start their backoff. The backoff of station C expires before the backoff of other stations, and station C transmits its RTS packet. When low priority stations hear the RTS, they freeze their backoff and reattempt the contention later.

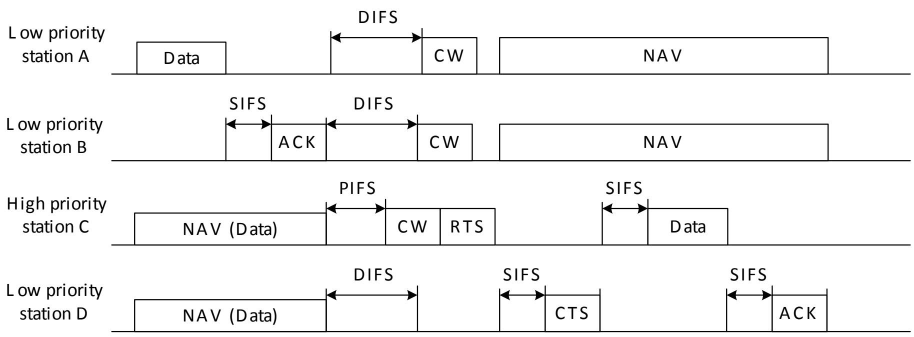

Figure 2-6. Service differentiation in DCF-PC.

This scheme is not sufficient to give higher chance of channel access to higher priority stations. The total time a station waits before wining access to the channel is the sum of the IFS and the random backoff time. However, higher priority stations may choose high backoff values. The higher priority station can still lose out contention to another station that has larger IFS but shorter random backoff value. Prioritization is reinforced by an extension to the backoff scheme of the IEEE 802.11 so that higher priority stations are assigned shorter backoff values.

HCF and IEEE 802.11e

Among all research activities on IEEE 802.11, the IEEE 802.11e [IEE 03] is the most promising framework and is expected to become a new industrial standard soon. IEEE 802.11e defines a new MAC sub-layer function called hybrid coordination function (HCF). HCF uses a contention-based channel access method, called Enhanced Distributed Channel Access (EDCA), which operates concurrently with a polling-based HCF-controlled channel access (HCCA) method. HCCA is an extension of the polling scheme initially introduced in the PCF and provides a centralized polling scheme to allocate guaranteed channel access to traffic flows based on their QoS requirements. As HCCA is based on a centralized mode, it is not suitable for ad hoc networks and therefore is out of the scope of this thesis.

Stations and access points that implement the HCF channel access scheme are called QoS-enabled stations (QSTA) and QoS-enabled Access Points (QAP) respectively. Similar to the original MAC, a CFP and a CP alternate over time, and time is divided into super-frames. The EDCA is the basic access scheme during the CP, while the HCCA is used in the CFP.

A fundamental difference between 802.11e and the legacy version is the concept of transmission opportunity (TXOP). A TXOP is a period of time during which a station has the right to use the wireless medium for transmission. A TXOP may be obtained during the contention period (CP) after winning the EDCA contention and it is called EDCA-TXOP, or polled by the QAP during the contention free period (CFP); in this case it is called HCCA-TXOP.

EDCA defines the concept of Access Categories (ACs), which can be considered as instances of the DCF access mechanism that provides support of prioritized channel access. Before entering the MAC sub-layer, each data packet received from the higher layer is assigned a specific user priority. At the MAC sub-layer, EDCA introduces four First-In-First-Out (FIFO) queues called access categories (ACs). Each data packet received from the higher layer is mapped into one of the four ACs according to its user priority. Each AC can be considered as an instance of the DCF with its own contention parameters (CW_{min}[#AC], CW_{max}[#AC], AIFS[#AC], and $TXOP_{limit}$[#AC]). (#AC = 0, 1, 2, 3 is the number of AC). Basically, the smaller the values of CW_{min}[#AC], CW_{max}[#AC], and AIFS[#AC], the shorter the channel access delay for the corresponding AC and the higher the priority for access to the medium. In EDCA, a new type of IFS is defined, the Arbitrary IFS (AIFS), in place of DIFS used in DCF. Each AIFS is an IFS interval with arbitrary length as follows:

$$AIFS[AC] = SIFS + AIFSN[AC] \times slot\ time \qquad (2\text{-}1)$$

Where AIFSN[#AC] is called the arbitration IFS number. After sensing the medium idle for a time interval of AIFS[#AC], each AC calculates its own random backoff time ($CW_{min}[\#AC] \leq$ backoff time $\leq CW_{max}[\#AC]$). The purpose of using different contention parameters for different queues is to give a low priority class a longer waiting time than a high-priority class, so the high-priority class is likely to access the medium earlier than the low-priority class. The EDCA access scheme is depicted in Figure 2-7 and Figure 2-8.

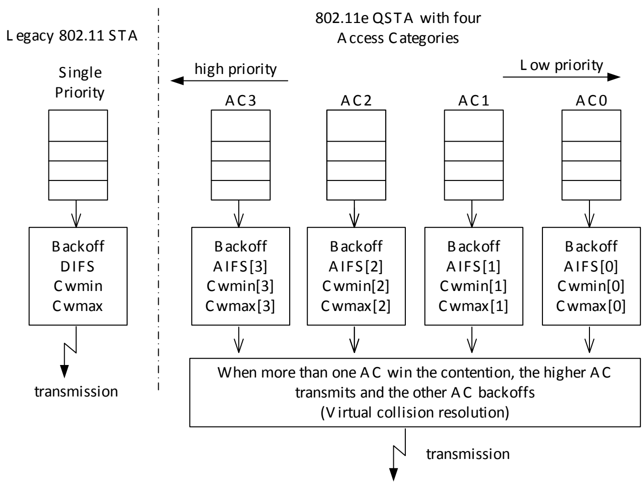

Figure 2-7. Queue model in IEEE 802.11e.

Figure 2 8 shows the channel access in EDCA with three AC. AC3, which has a higher priority, waits the channel to remain idle for AIFS[3] after which it starts its backoff timer. AC2 and AC1 must wait the channel to remain idle for AIFS[2] and AIFS[1] respectively before they start their backoff. Since AC3 has a shorter AIFS and a shorter CW its backoff expires before the backoff of other stations. At the backoff expiration, AC3 gets an EDCA-TXOP, while the other ACs freeze their backoff.

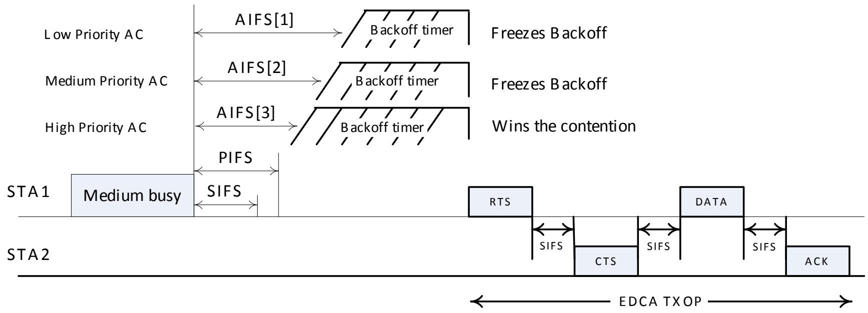

Figure 2-8. EDCA access scheme.

DPS protocol

DPS (Distributed Priority Scheduling) protocol [KAN 02-1, KAN 02-2] provides a framework for dynamic priority packet transmission scheduling in MANETs. DPS defines two mechanisms in order to provide QoS. The first scheme is the distributed priority scheduling; a technique that piggybacks the priority tag of a station’s head of line packet onto control and data transmitted packets. By monitoring transmitted packets, each station determines its priority level compared to other stations in its neighborhood. This priority level is used in the backoff calculation mechanism in order to approximate an idealized schedule in a broadcast region.

Each packet has an associated priority index, which is computed based on local information (e.g. deadline of the packet). The Head-of-line (HOL) packet of a station refers to the packet with the highest priority (lowest index) that is queued locally. The DPS protocol is based on the exchange of priority indexes of the current and HOL packets in a broadcast region. Each station maintains a local scheduling table, and on hearing new priority indexes, it adds them to its local table. This scheduling table is used to control the channel access.

When a source station transmits a DATA packet, its HOL packet information (the priority level of the HOL packet) is piggybacked on the transmitted DATA packet. This information is copied by the receiver onto the ACK packet it sends in response to the received DATA packet. Neighbor stations hearing the DATA and ACK packets retrieve the piggybacked information and update their scheduling table accordingly. When a station hears an ACK packet, it removes from its scheduling table any entry made earlier for the corresponding DATA packet.

The entries in the scheduling table are ordered according to their priority tag values. The rank of the station in the scheduling table (noted *r*) determines the priority of the station with respect to other stations in its neighborhood, and determines the backoff to be chosen by the station. The relationship between the rank of a station and its backoff timer is given by following (2-2).

$$backoff = \begin{cases} Uniform[0,(2^n CW_{\min})-1] & r=1, n<n_{\max} \\ \alpha \times CW_{\min} + Uniform[0,\gamma CW_{\min}-1] & r>1, n=0 \\ Uniform[0,(2^n \gamma CW_{\min})-1] & r>1, n\geq 1 \end{cases} \tag{2-2}$$

where CW_{min} is the minimum size of the contention window. n is the current number of transmission attempts. n_{max} is the maximum number of retransmissions. α and γ are constant parameters.

The second scheme is called multi-hop coordination. The latter is used in order to make up for delay that the packet has encountered on upstream stations by updating the priority index of the packet. When an intermediate station (the station has to forward the packet) receives a DATA packet, it receives its priority index piggyback. The station updates the priority index of the packet. Through the priority update scheme, if a packet suffers excessive delay at the upstream stations, downstream stations will increase the packet priority index so that the chance that the packet meets its end-to-end delay requirements is increased.

BB-DCF protocol

The BB-DCF (Black-Burst DCF) protocol [SK 96, SK 99] is an enhanced MAC protocol to transport real-time traffic over IEEE 802.11 wireless LANs. The protocol is based on two access mechanisms. The first mechanism is traffic differentiation. The protocol achieves service differentiation by assigning longer inter-frame spaces to data stations (i.e. stations injecting best effort traffic) than to real-time stations (i.e. stations injecting real-time traffic). The protocol defines three inter-frame spaces, denoted t_{short}, t_{med}, and t_{long}. The second mechanism is real-time access scheduling through jamming the wireless medium.

Access scheme for best-effort traffic: When a data station has a packet for transmission at time t, it transmits the packet immediately if the channel was idle in the interval $[t - t_{long}, t]$. Otherwise, it waits until the channel remains idle for t_{long}, and enters in backoff mode. Similarly, a station whose packet has experienced c collision waits until the channel is perceived idle for t_{long} and enters the backoff mode. The station initializes its backoff timer with a value *of* $t_{wslot} \times rand(f_{data}(c))$, where t_{wslot} is the duration of a slot, the function $rand(b)$ returns a random number between 0 and b-1, and $f_{data}(c)$ is given by equation 2-3.

$$f_{data}(c) = f_{data}(0) \times 2^c, c = 0,1,... \tag{2-3}$$

The timer counts down until the channel is perceived idle more than t_{long}, and the packet is transmitted as soon as the timer reaches 0. An acknowledgement scheme is used to indicate correctly received packets.

Access scheme for real-time traffic: Because of collisions, real-time stations may suffer excessive access delay before they can transmit their real-time packets. BB-DCF permits to limit the delay experienced by real-time packets. Real-time stations send jamming signals (called black bursts) in order to determine if they have the highest priority to transmit their real-time packets. The length of the black burst signal is proportional to the contention delay experienced by the station.

The access instant of a real-time station is defined as the time at which the station acquires access to the channel to transmit a packet. Each time a station has an access instant, it transmits its real-time data packet, and schedules its next transmission (next access instant) to occur t_{sch} later. Suppose now, that a real-time station has an access instant scheduled to occur at time t. If the channel was sensed idle in the interval [t - t_{med}, t] and remains idle during the following t_{obs} time (t_{obs} is equal to the maximum propagation delay), the station has an access instant at this later time (t + t_{obs}). Otherwise, the station waits until the channel remains idle for t_{med}, and enters in a black-burst contention period. That is, the station jams the channel with a number of black slots. The duration of the black burst is proportional to the time that the station has been waiting the channel to be idle. Suppose that the station has been waiting access to the channel for a period d, the duration of the black burst is calculated as follows (equation 2-4):

$$t_{BB} = t_{bslot} \times \lceil (d / t_{unit}) \rceil \tag{2-4}$$

Where t_{bslot} is the length of a black slot and t_{unit} is a system parameter. In order to know if any other real-time station has transmitted a longer black burst, the station senses the channel for t_{obs}. If the channel is sensed idle during t_{obs}, the station is alone transmitting a black burst, and so is allowed to transmit its data packet. Otherwise, that means that there is another station that is transmitting a longer black burst, implying that it would have been waiting longer to access the channel. In this case, the station must wait the channel to be idle for t_{med} and repeats the contention.

The black burst contention scheme is illustrated in Figure 2-9. Real-time stations *1* and *2* acquire access instances at times t_1 and t_2 respectively. At scheduled instants t_{sch1}, t_{sch2} stations *1* and *2* are not allowed to transmit their real-time packets because there is a data packet being transmitted. When the channel becomes idle, station *1* and station *2* enter the black burst contention period. Station *1* sends a black burst longer than the one sent by station 2, since station *1* has missed its transmission for a longer time. Station *1* transmits its queued data packet at time t_3, while station *2* repeats the black burst contention (at time t_4) at the end of transmission.

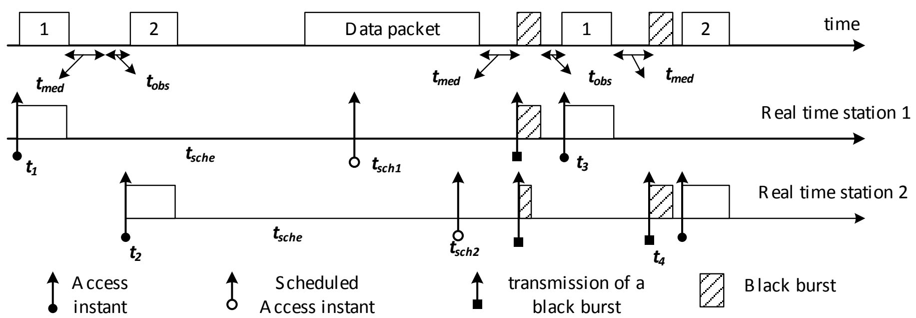

Figure 2-9. Channel access in BB-DCF.

ES-DCF and DB-DCF protocols

The ES-DCF (Elimination by Sieving DCF) and DB-DCF (Deadline Bursting DCF) [PAL 01, PDO 02] are two distributed collision resolution algorithms to provide timely delivery guarantees to different classes of real-time traffic. ES-DCF defines three channel access phases: Elimination, Channel acquisition, and Collision resolution. In the Elimination phase, a grade and a channel-free-wait-time (*CFWT*) are assigned to each station. The *CFWT* is calculated based on the deadline and priority of the packet to be transmitted. The closer the deadline of the packet, the smaller its *CFWT*, and the higher the grade. The grade of a real-time packet is improved as it remains in the queue for a longer time.

After waiting the channel to be idle for the assigned *CFWT*, the real-time station enters the Channel acquisition phase. In this phase, the station sends an RTS packet to the intended data receiver. Other real-time stations that have a larger *CFWT* defer their access upon hearing this transmission. These stations repeat their elimination phase at the end of the current transmission. If the RTS reaches successfully the destination, the transmitter receives a CTS packet from the receiver, and can begin the transmission of its data packet. Otherwise, the Collision resolution phase is initiated by sending a Black Burst (BB) signal. Unlike BB-DCF (where the length of the BB signal is proportional to the deadline of the real-time packet), the length of the BB signal is proportional to the station identifier (ID). After transmitting its black burst, the station turns around and listens the channel, if it hears a black burst of longer duration (transmitted by a station with higher ID number), it defers its impending transmission attempt. Otherwise, it repeats its channel acquisition phase, followed by a data transmission phase. Whenever a real-time station is forced to defer to another real-time station, the collision resolution phase is repeated after the channel becomes idle.

If a station experiences a collision in the Channel acquisition phase, it uses the smallest channel-free wait time (which is equal to PIFS), so it can pre-empt all other stations during collision resolution. The three phases of ES-DCF are illustrated in Figure 2-10. A real-time station has to wait for at least an amount of time equal to (*PIFS* + *slot_time*) before it can start the Elimination phase.

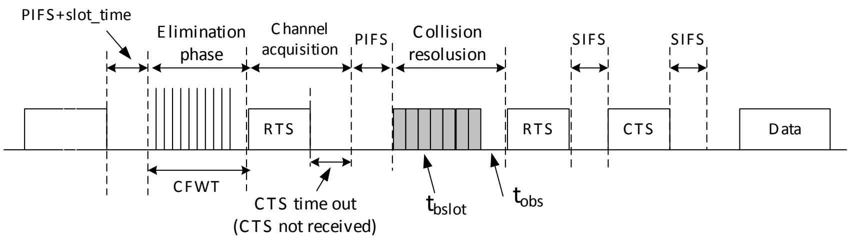

Figure 2-10. Channel access in ES-DCF.

The Deadline Bursting (DB-DCF) protocol is similar to ES-DCF with the exception that it uses the black-burst contention phase in order to resolve contention. At the beginning of a transmission cycle, a real-time station transmits its RTS packet if the channel is sensed idle for the duration of *PIFS* + *slot_time*. Otherwise, the station waits the channel to become free, and remains free for *PIFS* + *slot_time* duration, and then initiates its Black Burst (BB) contention phase. In the BB contention phase, each real-time station transmits a black burst packet of length proportional to the relative deadline of its packet. When the station finishes the transmission of its BB packet, it turns around and listens the channel for any longer-duration Black Bursts. If the station hears another BB, it defers its channel access, and repeats the BB contention phase later when the channel becomes free. This scheme is illustrated in Figure 2-11.

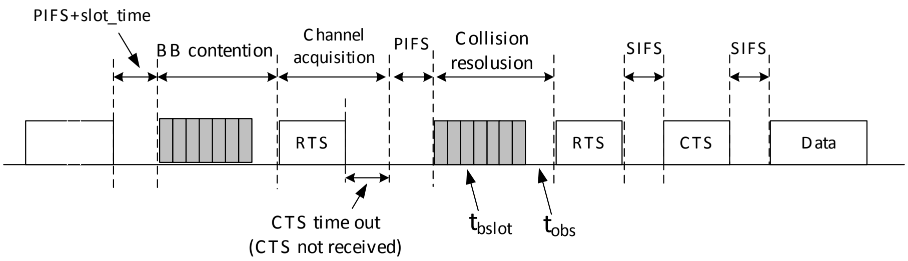

Figure 2-11. Channel access in DB-DCF.

Discussion

Contention MAC protocols with prioritization have been proposed in the aim of increasing the probability of successful packet transmission for real-time traffic sources through increasing the opportunity of contention for real-time traffic sources and penalizing data traffic sources.

Despite the huge work in this field, satisfactory solutions are still missing. Several researches show that, even with traffic prioritization extensions, contention-based MAC protocols are still inefficient, and cannot provide the needed QoS for real-time applications [RNT 03, MG 06]. The hidden and exposed terminal problems are still present, leading to an increase of the collision rate, and underutilization of the limited bandwidth, especially at high traffic load.

Constrained by the hidden and exposed terminal problems and traffic load, most of these prioritization MAC schemes provide only *soft* QoS, which means that the QoS is provided in a statistical manner. For example, higher priority packets will have better chance to access the channel than low priority packets on average over the long term, but the priority for a specific packet at a specific time cannot be guaranteed. Thus, soft QoS stands for QoS differentiation rather than QoS guarantees [BOU 06]. Indeed, to obtain a deterministic QoS, an allocation approach that provides the required guarantees is necessary.

2.4.2.2 Reservation-oriented Protocols

Reservation MAC protocols try to provide bandwidth reservation through giving a deterministic channel access for traffic sources with strict delay requirements. Thus, they seem to be a promising approach to provide QoS for real-time and multimedia applications.

In these protocols, the channel time is segmented into contiguous segments called super-frames. Each super-frame is composed of fixed length time-slots. The basis of these protocols is to give to each real-time traffic source a guaranteed periodic access to the wireless channel by reserving some slots of the super-frame to the traffic source. Once the reservation is done, the flow uses the reserved slots in subsequent super-frames without need for contention resolution. A particular aspect of these protocols is that the most of them require strict coordination between mobile stations in order to maintain coherent reservation over time. Examples of protocols in this category are FPRP, D-PRMA, SRMA/PA, MACA/PR and RTMAC.

Reservation protocols can be considered as extension of the TDMA (Time Division Multiple Access) to provide a spatial reuse of the available bandwidth. These protocols can be classified into two categories: synchronous and asynchronous protocols. Synchronous protocols require tight clock synchronization where all nodes are synchronized on a super-frame basis. As examples of this category of protocols, we can mention FPRP [ZC 98], D-PRMA [SHE 02], CATA [ZG 99], and R-CSMA [INW 04] protocols. Asynchronous protocols do not require clock synchronization; they rely on relative time information in order to coordinate reservations. The most representative schemes in this category are MACA/PR [LG 97] and RTMAC [MS 02, BMS 04, and MVS 04]. Before presenting these protocols, we introduce some basic concepts related to reservation MAC protocols.

2.4.2.2.1 Slot allocation conditions

In any reservation scheme, a node's use of a slot depends not only on the status of its one-hop neighbors use of this slot, its two-hop neighbors current use of this slot must be considered as well [BOU 06]. This constraint is imposed by the nature of the wireless medium and the hidden and exposed terminal problems.

A time-slot *t* is considered free to be allocated to send traffic from a node *x* to a node *y* if the following conditions are met:

1. Slot *t* is not scheduled for receiving or transmitting in neither node *x* nor *y*.
2. Slot *t* is not scheduled for receiving in any node *z* that is a one-hop neighbor of *x*.
3. Slot *t* is not scheduled for sending in any node *z* that is a one-hop neighbor of *y*.

These conditions should be met by each sender/receiver pair that has reserved a slot at the moment of reservation. In addition, all the other nodes which are not concerned by the reservation should be aware of reservations of their one-hop and two hop neighbors in order to not violate these conditions.

Fulfilling these conditions is considered as an important factor in the performance of reservation MAC protocols. While the first condition can be easily satisfied (as it depends only on local information), the second and third conditions constitute quite a problem. Both conditions require a strict coordination between one-hop and two-hop neighbor nodes. In order to avoid the violation of these conditions, some reservation protocols (such as MACA/PR, RTMAC, and R-CSMA) use reservation tables where each node keeps track of the slot status information of its one-hop and two-hop neighbors. In other protocols such as CATA, SRMA/PA, and D-PRMA, each node that reserved a slot sends periodically a signaling packet to prevent its neighbors from reserving its reserved slot.

2.4.2.2.2 Reservation protocols without synchronization

MACA/PR

MACA/PR (Multiple Access Collision Avoidance with Piggy-Backed Reservation) protocol [LG 97] is an extension of the IEEE 802.11 protocol to provide bandwidth guarantees to real-time traffic. In MACA/PR protocol, a station, which needs the establishment of a connection, sends a reservation request using contention. The reservation request specifies the reservation cycle and the amount of requested bandwidth. Protection of real-time packets from collision is ensured through maintaining reservation tables (RT) in neighboring stations. The reservation table includes for each reservation the reservation cycle and the amount of reserved bandwidth in each cycle.

Real-time stations in MACA/PR use the RTS/CTS/DATA/ACK handshake to establish reservations and resolve the hidden terminal problem. The first packet of a real-time flow sets up reservation and is transmitted after the RTS/CTS exchange. That is, the source station sends out an RTS packet, and waits a response from the receiver. On the reception of RTS packet, the receiver checks its reservation table and replies by sending a CTS packet. The source station sends its DATA packet, and schedules its transmission to occur after CYCLE time period. The control information of reservation (including the reservation cycle and the length of reserved slot) is piggybacked on the header of the transmitted DATA packet. On reception of the DATA packet, the receiver records the reservation in its reservation table, and confirms the reservation by sending an ACK packet. The control information of the reservation is extracted from the DATA packet, and piggybacked on the header of the ACK packet. Neighbors of the sender and receiver that hear the DATA and ACK packets record the reservation in their reservation tables and are not allowed to use the channel during the reserved slot.

The RTS and CTS control packets are needed only for the transmission of the first DATA packet. Subsequent DATA packets are transmitted cyclically on reserved slots without collision. For each cycle, the reservation is refreshed using the previous reservation information piggybacked on DATA and ACK packets. If during the connection the sender fails to receive ACK packets for a transmitted DATA packet, it concludes that the reservation has been lost. In this case, the sender cancels its reservation, and restarts the reservation process again through the RTS/CTS/DATA/ACK handshake. The reservation process is depicted in Figure 2-12.

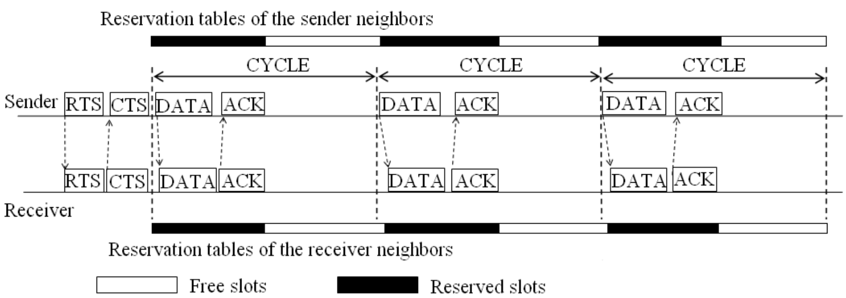

Figure 2-12. MACA/PR reservation process.

In order to avoid collision due to hidden terminals during reserved slots and maintain reservations coherence, neighbor stations are required to exchange their reservation tables periodically. This is a drawback for MANETs where the bandwidth is a limited resource. Moreover, errors may occur during the exchange of reservation tables, which leads to reservation missing.

RTMAC

RTMAC protocol (Real-Time Medium Access Control Protocol) is another extension of IEEE 802.11 DCF for bandwidth reservation in MANETs [MS 02, BMS 04]. Each station has its own super-frame which consists of a number of reservation-slots (resv-slots). A station that has real-time packets to transmit, reserves a block of consecutive resv-slots, which is called connection-slot, and uses the same connection-slot to transmit in successive super-frames. The reserved connection-slot is located on time axis using relative times of starting and ending times of the connection-slot. By this means of relative time of connection-slots, RTMAC eliminates the need for time synchronization. Each station maintains a reservation table that records for each reservation the starting and ending times of the reserved connection-slot.

To reserve a connection-slot, a station follows a three way handshake mechanism (*ResvRTS-ResvCTS-ResvACK*). First, the station waits for the channel to remain idle for DIFS period, and initiates a backoff timer. At the backoff expiration, the station transmits a Reservation request (*ResvRTS*) packet. The *ResvRTS* includes the relative time information of the connection-slot to be reserved (the start time of connection slot which is the number of resv-slots from the time of transmission of the *ResvRTS*; and the end time of the connection-slot which is the number of resv-slots from the transmission of the *ResvRTS*). Upon receiving the *ResvRTS*, the receiver extracts the relative time information from the *ResvRTS*, and converts it to absolute time by adding its current time maintained in its clock. The receiver checks then its reservation table to see whether it can receive in the specified connection-slot. If it can, it replies with a *ResvCTS* packet including the relative time information of the connection-slot to be reserved. Upon receiving the *ResvCTS*, the sender confirms the reservation by sending a *ResvACK* packet including the relative time information of the reserved connection-slot. Neighbors of the sender and receiver hearing the *ResvCTS* and *ResvACK* record the reservation in their reservation tables. Once the reservation is established, real-time packets are transmitted on the reserved connection-slot and are acknowledged by the receiver by Real-time ACK (RTACK) packets. The three-way handshake in RTMAC is illustrated in Figure 2-13.

When a real-time source finishes its transmission or if it detects a path breakage, it releases the reserved connection-slot by sending Reservation Release RTS (*ResvRelRTS*) packet. If the receiver receives the *ResvRelRTS*, it sends a *ResvRelCTS* packet. The purpose of the *ResvRelRTS* and *ResvRelCTS* packets is to request neighbors of the sender and receiver to release the reserved connection-slot.

Like MACA/PR protocol, RTMAC protocol has the advantage that it permits to reserve time slots of variable length. Its major drawback is the slots release mechanism. Reserved slots in RTMAC are released through the transmission of reservation release packets (i.e. ResvRelRTS, ResvRelCTS). As collisions may occur during reservation release packet transmission, some neighbors (of the sender or the receiver) are not notified of the reservation release, and needlessly maintain previous reservations.

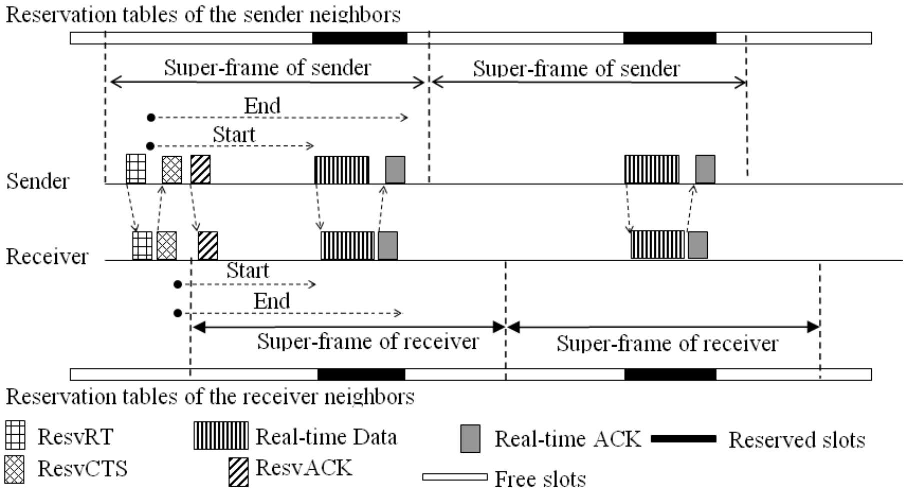

Figure 2-13. RTMAC reservation process.

Discussion

The major advantage of asynchronous reservation protocols is that they do not need global time synchronization. The second advantage is that each connection chooses the length of its reservation cycle, the think which is not available in synchronous protocols as will be shown later. One of their drawbacks is that small fragments (i.e., small portion of the wireless channel in the time scale) appear because nodes are allowed to reserve any portion of the wireless channel. These small fragments can be reserved only if they can fit the entire DATA/ACK exchange, and if they are free in each cycle. Consequently, it may happen that many of these free fragments not being used at all, leading to inefficient utilization of the limited bandwidth.

2.4.2.2.3 Reservation protocols with synchronization

With the growing development of GPS technology and the reduction of its cost, solutions for resource reservation based on time synchronization have become feasible for ad-hoc networks. The channel is viewed as a succession of super-frames, and each super-frame is composed of a fixed number of time-slots. An amount of the super-frame (or slot) is dedicated for the transmission of reservation requests and contention resolution. Several protocols were proposed in this category, they differ mainly in the frame structure and the adopted medium access mechanism.

D-PRMA

In [SHE 02], authors propose the D-PRMA protocol for resource reservation in Ad-hoc Networks. In D-PRMA protocol, each super-frame is composed of *S* slots, and each slot is composed of *m* mini-slots. Each mini-slot is composed of two control fields: RTS/BI field and CTS/BI field (Figure 2-14). These control fields are used for the purpose of time-slot reservation and hidden terminal problem avoidance.

A station, which has real-time packets to send, must first reserve a slot. When a station detects that channel is idle in the first mini-slot (mini-slot 0), it transmits an RTS packet during the RTS/BI control field. At the reception of the RTS, the destination station transmits a CTS packet during the CTS period. If the sender receives a CTS, it concludes that it has successfully reserved the current slot. If no station succeeds to make reservation in the first mini-slot, contention is repeated in a randomly chosen mini-slot. Contending stations repeat the reservation process until one of them succeeds in making reservation. Once allocated the same slot is reserved for reserving station in subsequent frames until the end of its transmission.

To give priority to stations which have real-time traffic over data traffic sources, only real-time traffic sources are allowed to transmit in the first mini-slot with a probability $p = 1$. The data source stations are allowed to transmit in the first mini-slot with a probability $p < 1$. Only real-time traffic sources are allowed to make periodic reservation of the same slot. Data sources are allowed to make reservation only in the current time-slot and are required to make reservation for each data packet.

In order to maintain reserved time-slots in subsequent super-frames, the sender and receiver of the reservation must transmit a Busy Indication (BI) signal in the RTS/BI and CTS/BI of the first mini-slot of reserved time-slots. By this means, neighbors of both the sender and receiver avoid contention in the reserved slots.

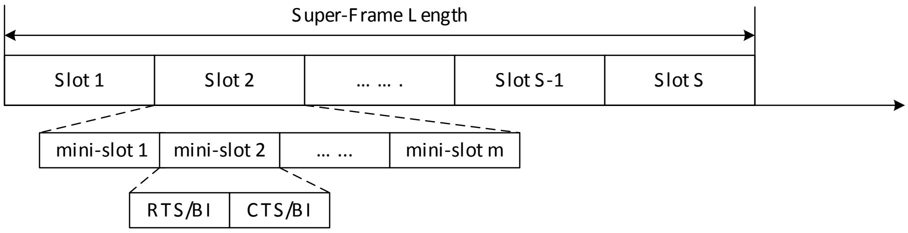

Figure 2-14. D-PRMA frame structure.

FPRP

The FPRP (Five Phase Reservation Protocol) [ZC 98] is a broadcast scheduling protocol for TDMA-based ad-hoc networks. Time-slot reservation is done in a distributed manner and based on contention. The super-frame structure of FPRP is shown in Figure 2-15. The super-frame is composed of Reservation Frame (RF) followed by several Information Frames (IF). Each RF is composed of *N* reservation slot (RS), and each IF is composed of *N* information slot (IS). In order to reserve an IS, the station must do a reservation during the corresponding RS. Each RS is composed of *M* reservation cycles (RC), and in each RC a five step reservation process is followed in order to do reservation in the current RS:

- Reservation Request: a station, which wants to set up a reservation, sends a Reservation Request (RR) during this phase.

- Collision report: if a collision is detected at the destination station during the Reservation Request, it transmits a collision report frame to inform the RR.

- Reservation confirmation: if the source station does not receive a collision report it concludes the RR frame was correctly received, in this case it transmits a Reservation confirmation frame to the destination.

- Reservation Acknowledgement: in this phase the destination acknowledges the RC frame by sending a Reservation Acknowledgement (RA) frame.

- Packing and eliminating (P/E): two packets are sent in this phase: packing packet and elimination packet. These two packets are used in order to resolve possible deadlocks between adjacent stations.

The advantage of FPRP is the support of multicast reservation. Its major drawback is the increase of reservation acceptance delay. When a station receives a reservation request during the information frame, it must wait until the next Reservation Frame (the termination of all information frames) before it can establish reservation. This may increase the reservation acceptance delay and by consequence the increase of end-to-end delay of real-time packets. Another drawback is the waste of RC slots after a station has done successful reservation. When a station successfully reserves a slot during a Reservation Cycle (RC), the remaining Reservation Cycles are unused for reservation request since the RC corresponds to only one Information Slot (IS).

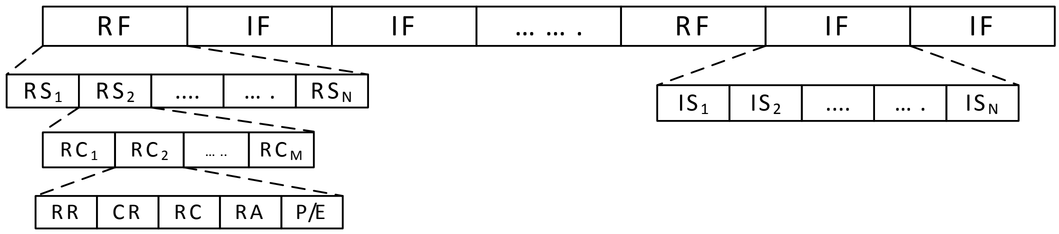

Figure 2-15. FPRP frame structure.

CATA

CATA (Collision Avoidance Time Allocation) protocol [ZG 99] is designed for the support of unicast, multicast and broadcast reservation. CATA divides time into equal size frames, and each frame is composed of *S* slots. Each slot is composed of five mini-slots. The first four mini-slots are used to transmit control packets and are called Control Mini-Slots (CMS1, CMS2, CMS3, and CMS4). The last mini-slot is the Data mini-slot (DMS), and is used for the transmission of data. Each station receiving data during the DMS mini-slot of the current slot sends a Slot Reservation (SR) packet during the CMS1 of the slot. The frame structure in CATA is illustrated in Figure 2-16.

For a given source station *n* and time slot *t*, the CATA protocol works as follows. Regardless packet type, *n* must first determine whether the current slot has been previously reserved or not. To reserve a slot, all stations that previously received data in slot *t* send a slot reservation (SR) control packet in CMS1. In addition, each source station that needs to maintain a reservation sends an RTS and not-to-send (NTS) control packets in CMS2 and CMS4, respectively. If no SR packet is detected in CMS1, then source station *n* contends for slot *t* by sending its own RTS in CMS2. Reception of a unicast RTS causes a station to respond with corresponding CTS in CMS3, and source station *n* can transmit its data packet in the subsequent DMS. Reception of a multicast or broadcast RTS or detection of a clear channel in CMS2 causes a station to remain silent during CMS3 and CMS4; otherwise, it sends a NTS in CMS4 to indicate a potential problem for local multicast or broadcast transmissions. Detection of a clear channel in CMS4, allows source station *n* to transmit a multicast or broadcast packet in the subsequent DMS. Any unsuccessful attempt to use a slot in this manner is managed by a backoff scheme.

The advantage of CATA against the other protocols is its simplicity and support of unicast, broadcast and multicast reservation simultaneously. Its major drawback is the waste of bandwidth due to control traffic; CATA allocates four control mini-slots on each slot in order to maintain coherent reservations.

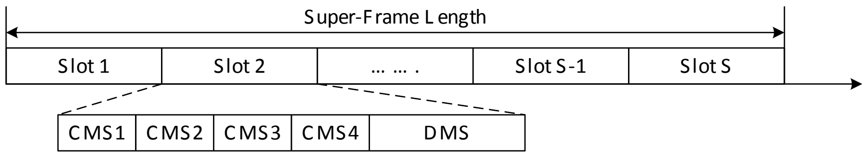

Figure 2-16. CATA frame structure.

SRMA/PA

The principle of SRMA/PA (Soft Reservation Multiple Access with Priority Assignment) protocol [CKC 03] is similar to CATA and D-PRMA protocols. In SRMA/PA, each slot is composed of six fields: SYN, Soft Reservation (SR), Reservation Request, Reservation Confirm (RC), Data Sending (DS), and Acknowledgment (ACK). The SYN field is used for synchronization. The SR, RR, RC, and ACK fields are used for the transmission of reservation control packets. The DS field is used for Data packets transmission (Figure 2-17), and ACK field for acknowledgements.

The advantage of SRMA/PA is its simplicity. Its major drawback is that only the sender of a reservation is able to notify its reservation to its neighbors through the Soft Reservation (SR). Neighbors of the receiver have no way to be aware about the reservation. These neighbors may attempt to reserve the same slot, and collision happens at the receiver. Another drawback of SRMA/PA is the waste of bandwidth consumed by unused control packets. The RR and RC mini-slots are maintained in the slot, even though they are not used for Data transmission or collision avoidance.

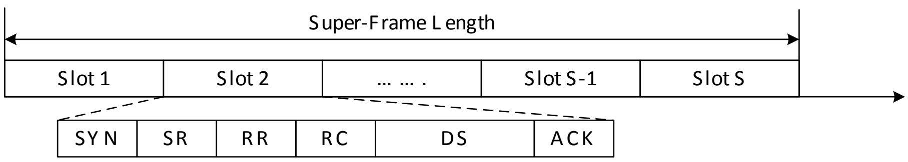

Figure 2-17. SRMA/PA frame structure.

R-CSMA

R-CSMA (Reservation CSMA) protocol [INW 04] is a distributed MAC reservation protocol. Time in R-CSMA is segmented into contiguous frames, and each frame is composed of a contention period (CP) and a set of *N* TDMA slots (Figure 2-18). R-CSMA is based on three-way handshake in making slot reservation. The goal of the three-way handshake is to fix the hidden terminal problem by announcing the reserved slot to all neighbor stations in the transmission range of the sender and the receiver. Each station maintains a reservation table describing the state of each time-slot “reserved” or “available”.

R-CSMA introduces three control frames to establish reservations: RFS (Request for Slot Reservation), RAC (Reservation Acknowledgement), and RAN (Reservation Announcement). A station, which needs to establish a reservation for a real-time traffic, negotiates the reservation with the intended receiver. The sender transmits a RFS frame to the receiver during the CP. The RFS frame includes the set of available slots from the sender standpoint. If the receiver receives correctly the RFS frame, it looks for a common free time-slot with the sender in its reservation table. If there is a common idle slot, the receiver grants the reservation by responding with a RAC (Reservation Acknowledgement) frame indicating which slot will be reserved for the real-time traffic. On the reception of RAC packet, the sender sends a RAN (Reservation Announcement) packet. This packet also includes the first data to be transmitted. The RAN frame is used to inform neighbors of the sender about the reservation. Neighbors of the sender and receiver record the reservation thus preventing any collision during reserved slots. Once the reservation established successfully, the sender starts using the reserved slot in future frame cycles without any collision until the end of transmission of the real-time traffic.

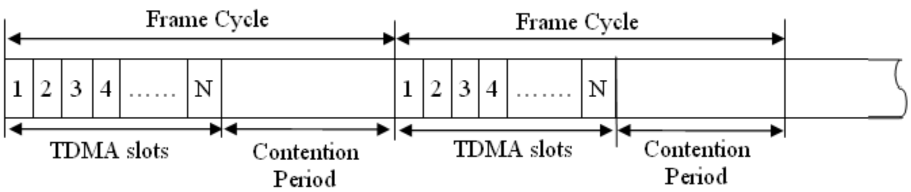

Figure 2-18. R-CSMA frame structure.

DARE

Most of protocols presented in this section focus only the point-to-point reservation problem. Only a few work addressed the problems related to the establishment of end-to-end reservations along a path. In [CAR 04, CAR 05, CAR 06], authors proposed the DARE protocol to deal with this problem. They assumed that there is a routing protocol providing the MAC sub-layer with routes between the source and the destination. To establish an end-to-end reservation, the source sends a Request-To-Reserve (RTR) message including the periodicity and duration of the reservation. Each intermediate node along the path which receives the RTR forwards the RTR to the next-hop if it can accept the reservation request. When the destination receives the RTR it generates a Clear-To-Reserve (CTR) along the reverse path to indicate to all intermediate nodes that the reservation request is accepted.

However, despite the routing protocol is involved, it is implicated only during the route discovery phase at the source, and the end-to-end reservation process is handled by the MAC sub-layer.

Despite it enables end-to-end reservation establishment, DARE protocol does not ensure consistency of reservations as there is no guarantee that reservation control packets (RTR and CTR) are received by all one-hop neighbors of intermediate nodes. Reservation breakage may occur even in a low mobility network.

Other approaches in the same category differ in the handshaking procedures and the arrangement of reservation phases with regard to the data transmission phases, such as [MG 06, JOE 05].

2.5 Summary

Due to the variable nature of the wireless channel, nodes mobility, and limited bandwidth in wireless ad-hoc networks, providing QoS over this type of networks is very challenging. Multimedia applications such as voice require some strict guarantees in terms of bandwidth requirements and transmission delay. Since MAC protocols directly control the sharing, and therefore the utilization the limited bandwidth resource, their role is critical to the overall QoS provided by the network. MAC protocols must grant channel to nodes in such a manner that increases bandwidth utilization, and minimizes collisions and channel access delay.

In this chapter, the major issues involved in the design of MAC protocols for Ad hoc Wireless Networks were identified. A classification of existing MAC protocols was provided, with presentation of the major MAC protocols that exist for Wireless Ad hoc Networks.

As has been discussed in the chapter, the design of MAC protocols for Ad Hoc Wireless Networks that provide QoS guarantees is very challenging. The popular IEEE 802.11 DCF scheme does not include any explicit quality of service (QoS) support [BOU 06]. All packets are treated and served in exactly the same manner. Thus, the packet delay, the dropping rate, and the offered session throughput cannot be predicted because they depend on the network conditions such as traffic load and mobility. The lack of predictability and controllability largely limit the DCF scheme in Ad hoc Networks to best effort applications only.

In recent years, numerous research efforts have been devoted to devise new MAC schemes, which we call QoS MAC protocols. The first category of these protocols consists in making QoS extensions to the legacy DCF scheme through service differentiation. Prioritization MAC protocols (such as IEEE 802.11e, DSP, ES-DCF, and BB-DCF) allow mobile terminals to access the wireless medium in a differentiated and controllable manner, in order to satisfy the QoS requirements of high priority flows. The IEEE 802.11e EDCF standard is the most representative of these protocols.

Experimental results [RNT 03] show that service differentiation enhances significantly the performance of the legacy IEEE 802.11. However, in spite of these enhancements, most of prioritization schemes brought to the IEEE 802.11 provide only a *soft* QoS. Indeed, to obtain a deterministic QoS, an allocation approach is necessary.

The reservation-based MAC protocols aim at providing QoS guarantees through ensuring a deterministic access to the wireless channel. Reservation-based access protocols include some features which are interesting for real-time and multimedia traffics such as low collision rate, low access delay, and the low impact of traffic load. We distinguished asynchronous and synchronous reservation protocols according to whether time synchronization is required or not, respectively.

Despite the advantage of the no need of global synchronization, asynchronous reservation protocols are inefficient in terms of bandwidth utilization due to the unusable small time fragments that appear in the wireless channel (cf. 2.4.2.2.2). This problem does not exist in synchronous reservation protocols because nodes are aligned on the slot and super-frame basis.

Synchronous reservation protocols require global synchronization where all nodes are aligned on the slot and super-frame basis. Protocols like CATA, D-PRMA, and SRMA/PA, control contention within a time-slot. Each time-slot is composed of a control part, and data transmission part. The control part is composed of a number of control mini-slots, while the data part contains generally one data mini-slot. Thus, the control mini-slots are dedicated to the slot, and are used to reserve the slot to which they are dedicated following some handshaking scheme. This solution has the drawback that the control mini-slots allow to reserve only one slot at a time. A node which wants to reserve several slots is required to compete and apply the handshaking scheme on each one of the slots that want to be reserved. Moreover, as the sender has no information about slots availability at the receiver, it may be required to compete several times before finding a slot which is available for reception at the receiver. Finally, the solution may suffer a waste of bandwidth if the number of control mini-slots per slots is high.

In other protocols (like R-CSMA, DARE, and RTMAC), a node has the possibility to reserve more than one slot in one handshake if needed. Indeed, unlike CATA and D-PRMA, where the sender has no information about slot availability at the receiver, the sender and receiver here, negotiate the set of common available slots that can be reserved. This is achieved through a two-way or three-way handshake during which the sender and receiver exchange their local information about the status of slots. One advantage of these schemes is that the reservation process is expected to be faster, especially when a traffic source needs to reserve several slots, because the source is not required to repeat the handshake scheme to reserve each one of the slots.

Although distributed reservation access schemes claim the ability to provide delay bound service to real-time traffic, several challenges face the deployment of this category of protocols. These challenges can be summarized in the following points:

- The first issue is how to ensure consistency of reservations. Most of reservation protocols implicitly assume reliable wireless links and perfect information exchange. However, this assumption is impractical in real world. Indeed, the ability of a reservation MAC protocol to provide a collision-free schedule depends on the ability of nodes to inform their neighbors about their current reservation, that is, to prevent neighbor nodes from reserving already reserved slots. Protocols like CATA, SRMA/PA, and D-PRMA use explicit signaling packets in order to jam any reservation request aiming to reserve currently reserved slots. However, this solution is inefficient in terms of control overhead and energy consumption. The other protocols (such as R-CSMA, RTMAC, MACA/PR, and DARE) do not use periodic signaling packets. Instead, they use reservation tables where reservations of neighbor nodes are recorded at the reception of reservation requests and reservation confirmations. However, the consistency of reservation tables are widely affected by the reliability of exchange of reservation control packets between neighbor nodes. Thus, there are situations where some neighbors of a node cannot be aware of a reservation established by the node because of collision of reservation packets. These neighbors may cause interference during reserved slots and violate the conditions of collision-free reservation explained in section 2.4.2.2.1.

- The second open issue with reservation protocols is related to how to set parameters of the protocol for the given network conditions to satisfy the required QoS. Thus, it is required to determine what is the appropriate super-frame length, the slot length, and contention period to resolve the contention. In addition, these protocols allocate reservation slots to resolve contentions and data information slots to transmit data packets. Thus, it is important to determine at what frequency the contention should be run. All these parameters have an important impact on the access delay of real-time traffic sources, and consequently on their QoS.

- The third issue is the suitability of these protocols for a particular traffic use. Most of these protocols are generic solutions as they don't take into account the specific characteristics of multimedia traffic types such as voice or video. For example, voice traffics are characterized by their ON/OFF characteristic due to the VAD (Voice Activity Detection). When a network node does not have traffic to send in its assigned slot, the time slot is wasted in these schemes. A mechanism that conserves the channel bandwidth is needed in such scenarios.

- Another issue is mobility handling. While acceptable delay can be expected at low mobility, reservation MAC protocols may be unstable under high mobility conditions. Thus, in a mobile environment, due to node mobility and potential change of topology, nodes may enter in the transmission range of each other resulting in collision during reserved slots. If no mechanism to handle these situations is used, collisions in one super-frame are automatically repeated in the next super-frames, resulting in frequent collisions and packet dropping. Except DARE, no one of the reservation protocols cited in this chapter considers performance degradation due to mobility of nodes. Thus, in a mobile environment, these protocols do not have any mechanism to converge on a new schedule, and as a result lose their ability to deliver packets with a delay guarantee.

The next step of our work consists of proposing a protocol that aims at solving some of these issues. In this protocol, we twin several mechanisms in order to ensure consistency of reservations, and reduce the performance degradation due to mobility of nodes. The proposed protocol is extended in order to take into account the characteristics of voice traffic.

Chapter 3: A solution for inconsistency of reservations

3.1 Introduction

The strength of a reservation protocol lies in its ability to provide collision-free transmission for traffic flows as long as possible. In order to guarantee such collision-free transmission, all nodes of the network should meet some rigorous slot allocation conditions (cf. section 2.4.2.2.1):

- The slot to be reserved should not be reserved neither for transmission nor for reception at both the sender and receiver;

- The slot should not be reserved for transmission by any one-hop neighbor of the receiver;

- The slot should not be scheduled for reception at any one-hop neighbor of the sender.

Violation of these conditions leads to a drastic consequences on the performance of the protocol in terms of packet collision, and consequently to excessive packet dropping rate. Thus, satisfying these conditions is a requirement in order to reduce the probability of collision, and dropping rate.

While the first condition is trivial, the two other conditions are more difficult to satisfy. Indeed, these conditions can be satisfied only if a tight coordination between nodes is provided. Therefore, each node of the network should be made aware of the slots reserved by its one-hop and two-hop neighbors. This coordination function requires the exchange of control messages that include information about slots reservation, and that are destined to inform all neighbor nodes about the current reservation.

In some reservation protocols (such as CATA and SRMA/PA), this tight coordination is achieved at the expense of excessive control overhead (cf. Chapter 2). In other protocols (such as R-CSMA and RTMAC), coordination is not always guaranteed due to possible collision of reservation control messages. There are still some situations where nodes are not able to record reservations of their neighbors, and consequently the conditions of collision-free slot reservation are not met.

In this chapter, we propose a reservation protocol aiming at reinforcing the coordination between nodes in order to achieve collision-free schedules, while keeping the overhead reasonable.

3.2 Inconsistency of reservations: causes and consequences

A basic issue in reservation MAC protocols is how to inform neighbor nodes about a reservation, and preventing them from reserving already reserved slots. One solution to do this, consists of sending periodically signaling messages on the reserved slot to indicate that the slot is reserved (such is the case with CATA, D-PRMA, and SRMA/PA). This solution is inefficient due to the high signaling overhead and power consumption that it incurs.

An alternative solution with low signaling overhead consists in sending reservation messages only at the reservation request phase, and using reservation tables. Neighbor nodes that hear reservation messages, record the reservation in their reservation tables. Hence, there is no need to send signaling messages again once reservations are recorded.

In reservation protocols that use this solution, reservation control messages include information about the future reservation such as periodicity of reservation and which slots will be reserved, so that the reservation can be recorded in reservation tables of neighbor nodes. Thus, these protocols should ensure that reservations established by a node are reported to all its neighbors in order to provide collision-free transmission. However, reservation conflicts may occur if a node is not able to receive a reservation packet transmitted by one of its neighbors, and consequently, cannot record the reservation. We call this problem the *inconsistency of reservations* problem.

In order to illustrate the importance of this problem, we take as example a three-way handshake reservation scheme (such as R-CSMA [INW 04], DARE [CAR 04-05-06], or RTMAC [MS 02, BMS 04]). In such schemes, a reservation is initiated by the sender through sending a reservation request (*ResvRTS*) message including the slots available for transmission from its point of view. When the receiver receives the *ResvRTS*, it replies with a *ResvCTS* specifying the slots which will be reserved. Neighbors of the receiver record the reservation and avoid reserving the reserved slot for transmission. Then, the sender informs its neighbors about the reservation through sending a reservation confirmation (*ResvConfirm*) message.

For the reservation, in order to be taken into account, all one-hop neighbors of the receiver must receive the *ResvCTS* collision-free, and all neighbors the sender must receive the *ResvConfirm* correctly. However, we identified some situations where some nodes cannot receive correctly reservations packets sent by their neighbors. These situations are due to collision of packets (*ResvCTS* or *ResvConfirm*) at neighbor nodes (scenarios of Figure 3-1 and Figure 3-2), or because two neighbor nodes are transmitting control packets simultaneously (scenarios of Figure 3-3 and Figure 3-4).

In order to illustrate these conflicts in detail let us consider scenarios of Figure 3-1, Figure 3-2, Figure 3-3, and Figure 3-4. In Figure 3-1, nodes E1 and E2 establish reservations on slots S1 and S2 with R1 and R2 respectively. If E1 and E2 transmit their requests at the same time, R1 and R2 also transmit their reservation responses at the same time. *ResvCTS* sent by R1 and R2 will collide at E3 and both reservations will not be recorded by E3. E3 may later attempt to reserve slot S1 or S2 with R3 which is out of the transmission range of R1 and R2. Since S1 and S2 were reserved by R1 and R2 for reception, collisions occur during these slots at R1 or R2. We call this scenario the ***unheard reservation scenario 1***.

The same problem rises in Figure 3-2 where the reservation confirmation sent by senders E1 and E2 (which have reserved slots S1 and S2) collide at node R3. R3 may accept a reservation request from E3 on slots S1 or S2, and collisions will occur on the reserved slot at R3. We denote this scenario the ***unheard reservation scenario 2***.

The two other scenarios where a node cannot hear reservation packets of its neighbors are situations when two neighbor nodes are transmitting at the same time. These two scenarios are shown in Figure

3-3 and Figure 3-4.

In Figure 3-3, we consider that nodes E1 and E2 simultaneously established reservations with R1 and R2 on slots S1 and S2 respectively. Since E1 and E2 send their reservation confirmation (*ResvConfirm*) at the same time they cannot receive the *ResvConfirm* sent by each other. E1 may accept a reservation request from E4 on slot S2 and collision will occur at E1 on slot S2 because S2 is reserved by E2 for transmission. We call this scenario the ***reservation deadlock scenario 1***.

The last scenario is similar to the previous one. In Figure 3-4, suppose that nodes E1 and E2 simultaneously reserved slots S1 and S2 with R1 and R2 respectively. Since R1 and R2 send their reservation response (*ResvCTS*) at the same time they cannot receive the reservation response sent by each other. R1 may reserve slot S2 for transmission with R4 and collision will occur at R2 on slot S2. Similarly, node R2 may reserve slot S1 for transmission to R3. Since R1 have previously reserved slot S1 for reception, collision occurs on slot S1 at R1, and R1 losses its reservation. We call this scenario the ***reservation deadlock scenario 2***.

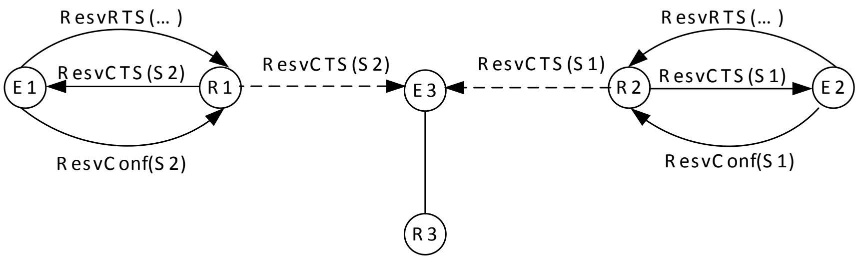

Figure 3-1. Unheard control packet problem scenario 1.

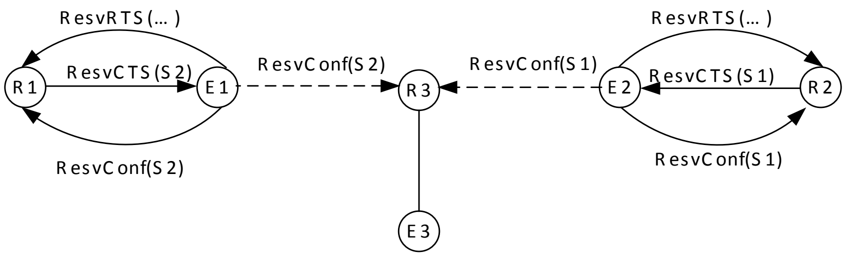

Figure 3-2. Unheard control packet problem scenario 2.

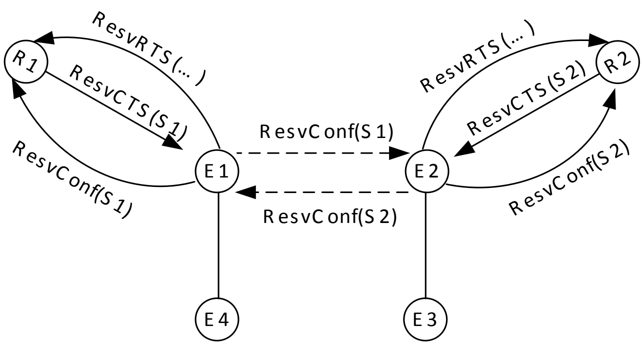

Figure 3-3. Deadlock problem scenario 1.

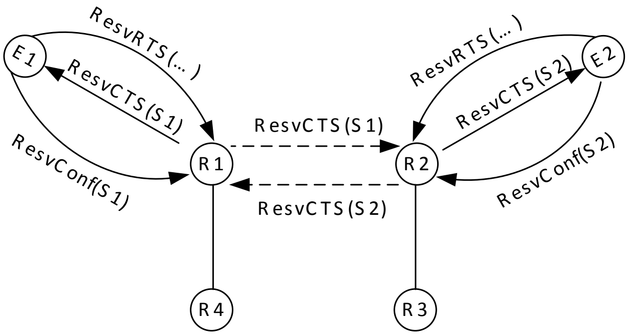

Figure 3-4. Deadlock problem scenario 2.

These scenarios can be generalized to any number of nodes scenarios.

These situations may lead to several conflicts that have drastic consequences on the overall performance of the reservation protocol. Thus, several questions should be answered:

- What is the best handshake to establish reservations?
- How to prevent collisions during the handshake?
- How to ensure that each reservation is taken into account by each neighbor?

In the rest of this chapter, we propose a protocol to answer these questions and eliminate some ones of the above discussed issues.

3.3 A protocol for reservation consistency

3.3.1 Design objectives

We showed that in order to be efficient, a reservation scheme should first of all avoid the different reservation conflicts and ensure consistency of reservations. In this direction, the proposed protocol aims to reduce the degradation that arises due to these conflicts. Our protocol is based on better coordination and cooperation between nodes in order to ensure consistency of reservations. The basic idea of this reservation scheme is to ensure that each reservation control packet transmitted by a node is successfully received by all its neighbors. While resolving the reservation conflicts, the protocol should keep the control overhead at low level.

3.3.2 Assumptions

We make the following assumptions regarding the networking context in which our protocol is expected to operate:

- The network is composed of n nodes that keep global time synchronization.

- Each node has a unique identifier.
- A link between two nodes pair is noiseless and symmetrical channel.
- No capture is considered, i.e., when multiple packets arrive to a node, collision is assumed and all of them are destroyed.
- The network topology is considered slowly changing relatively to the time required to compute a new schedule. Thus, nodes may move, but their speed is slow compared to the number of times a transmission schedule may be used. Thus, when a schedule is computed, it can be used for some time before a topology change requires a new schedule to be computed. Other mobility assumptions and their impact on the performance of the protocol are discussed later in chapter 7.
- The network is supposed to support a predefined real-time traffic with specific traffic specification and bandwidth requirements.

The connectivity between nodes forming the network is represented by an $n \times n$ symmetric one-hop connectivity matrix $Neigh_1$ given by:

$$(Neigh_1)_{ij} = \begin{cases} 1, & if\ nodes\ \ i\ and\ j\ are\ neighbors \\ 0, & otherwise \end{cases}$$

3.3.3 Super-frame design

The wireless channel in the time scale is divided into super-frames of fixed length. As our protocol is primarily designed to support real-time traffic, the super-frame should be structured in such a way that takes into account the characteristics of the targeted type of traffic (voice, video ... etc).

The super-frame length is set to the inter-packet arrival time of the considered real-time application. As shown in Figure 3-5, which represents the super-frame structure, each super-frame in our protocol is composed of a SYNC slot, followed by a Reservation Sub-Frame (*RSF*), followed by a Data sub-frame composed of *S* data slots. The SYNC slot is used for synchronization. All nodes are synchronized on the super-frame basis. A possible solution for synchronization consists in using the GPS that provides a global synchronization for all nodes.

Data slots are used to carry real-time and data packets, and their length (L_{slot}) is set to the transmission time of one real-time packet including the different layers (RTP, TCP/UDP, IP, and MAC headers) overheads. The choice of the slot length is discussed in Chapter 5.

The ACK mini-slot is used to acknowledge successful (or failure) reception of real-time and data packets through the transmission of ACK (or NACK) frame. The *RSF* is composed of *R* Collision Resolution Slots (*CRS*) used for reservation requests and reservation releases requests. Each *CRS* is composed of five control mini-slots used to reserve slots, and avoid collision of control packets and reservations conflicts. A guard time corresponding to the round trip propagation delay is inserted at the end of each control or data slot.

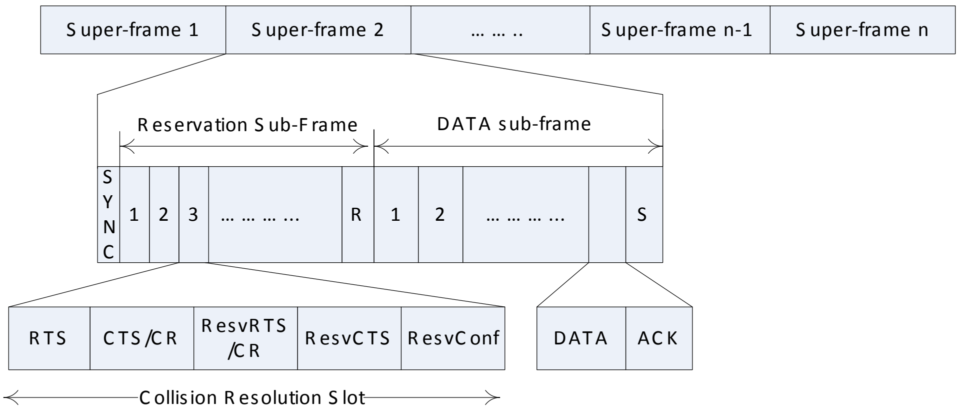

Figure 3-5. Super-frame structure of our protocol.

3.3.4 Slot Status

In order to maintain consistency of reservations with neighbor nodes, each node maintains a Slot State Table (SST). This SST reflects the actual status of slots in the super-frame, and is updated each time a slot is reserved locally or by neighbor nodes. In addition, it should be updated when reservations are lost or released as described later in Chapter 7.

In order to take into account the limited memory and processing capacity of mobile nodes, only the minimum information needed for reservation consistency maintenance should be recorded in the SST. The SST contains two fields: the "*slot number*" field and the "*slot status*" field which represents the current status of the slot. The SST may include other fields such as the "*traffic type*" field which represents the traffic type of the traffic for which the slot is reserved.

The values of the slot status field reflect the status of the slot, and the status of reserving nodes during the life time of the connection. In this chapter, we consider a simple representation of the slot status field where a slot is allocated to a traffic flow for all the duration of the connection. In this representation, we do not consider any difference between real-time and data traffic sources in slot reservation. Another representation of the slot status field is considered in chapter 5 in order to take into account the characteristics of real-time traffic sources and give them more priority than data sources.

For a node *i*, a slot may be in one of the following statuses:

- *Available for transmission and reception (ATR)*: no neighbor has reserved the slot neither for transmission nor for reception, and node *i* can reserve it for transmission or reception.

- *Reserved for transmission (RT):* one or more neighbors have reserved the slot for packet transmission. Node *i* can reserve the slot only for reception.

- *Reserved for reception (RR):* one or more neighbors have reserved the slot for packet reception. Node *i* can reserve the slot only for transmission.

- *Reserved for transmission and reception (RTR):* the slot is reserved for transmission by one or more neighbors and for reception by other neighbor nodes. Node *i* cannot reserve this slot neither for transmission nor for reception.

Another solution to record reservations of neighbor nodes is to use slot state variables instead of slot state tables, where each state variable represents the set of slots in a particular state. Thus, we need to define the two following state variables:

- *Reserved for Transmission slots set (RT).*

- *Reserved for reception slots set (RR).*

A slot which is not neither in the *RT* set nor in the *RR* set, is considered available for both transmission and reception. A slot which is present in both sets is considered non-available neither for transmission nor for reception.

Let RT_i and RR_i denote the slot state variables of node *i*. Initially, all data slots are considered available for both transmission and reception for all nodes, i.e., $\forall i, RT_i = RR_i = \phi$.

At any instant, the schedules of nodes should respect the following conditions:

- $\forall i, \forall j \in (Neigh_1)_{ij}, \forall t \in RT_i \Rightarrow t \notin RR_j$
- $\forall i, \forall j \in (Neigh_1)_{ij}, \forall t \in RR_i \Rightarrow t \notin RT_j$

In addition to the slots state, each node should record information about the flows crossing such a node. Thus, we define the *flow list table (FLT)* which records for each traffic flow that crosses the node, the *flow identifier (fid)*, the slot which is reserved to the traffic flow, and the identifier of the node for which the slot is reserved.

Figure 3-6 shows an ad-hoc network composed of 12 nodes with two voice traffic flows *f1* between nodes *N1* and *N4*, and *f2* between *N5* and *N8*. The figure shows also the slot which is reserved to each traffic flow on each one of the point-to-point links along the respective path, the slot state variables, and the *FLT* of each node.

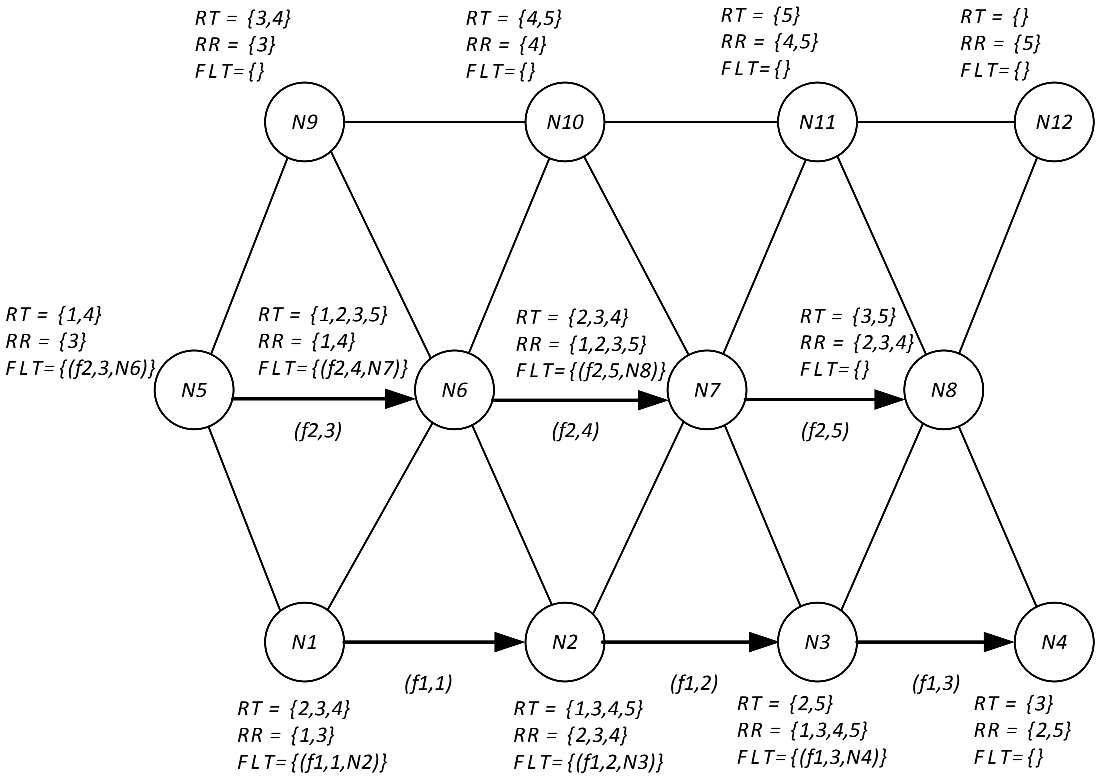

Figure 3-6. Illustration of the concept of slot state variables and flow list table on a network of 12 nodes. Arrows represent traffic flows, and the caption above each arrow represents the flow identifier and the slot reserved to the traffic flow.

3.3.5 Basic reservation scheme

In this section, we present the handshake scheme that allows nodes to establish consistent reservations, and avoid reservation conflicts. Our protocol tries to avoid these conflicts through better coordination between nodes. Indeed, the handshake scheme of our protocol consists in ensuring that a reservation is confirmed only if it is recorded in the *SSTs* of all the neighbors of both the sender and receiver. Thus, any unheard reservation due to collision of reservation control packets during the reservation phase is considered invalid.

Reservation of time-slots in our protocol can be summarized in the following steps:

1. Floor acquisition at the sender: This step aims at determining if there is another node two-hops away from the sender, which is contending simultaneously for reservation in the current *CRS*. Thus, a node which needs to establish a point-to-point reservation, first executes one of the contention resolution algorithms described in chapter 5 during the *RSF*. Once the sender has acquired the transmission right in a *CRS*, it sends an *RTS* packet in the 1st control mini-slot, and waits for a Collision Report from one of its neighbors or *CTS* from the intended receiver.

2. Floor acquisition at the receiver / Collision report: This step is considered as a replay to the *RTS* sent by the sender on the 1st control mini-slot. For the receiver, its aim is to answer the sender, and at the same time, to assess whether all neighbors of the receiver are ready to receive correctly a reservation control packet from the receiver in the current *CRS*. Thus, when the receiver receives correctly the *RTS* it replies with *CTS* in the 2nd mini-slot and waits for a *ResvRTS* from the sender or *Collision Report (CR)* packet from one of its neighbors in the 3rd mini-slot. For the other neighbors of the sender, it is used to notify any collision during the 1st control mini-slot. Thus, if a collision of the *RTS* occurs at any neighbor of the sender, the neighbor sends a *CR* packet in the 2nd mini-slot to indicate that more than one node are trying to establish reservation at the same time. In this case, the sender cancels its reservation request and restarts the collision resolution process. This avoids the situation where some neighbors of the sender don't receive the reservation request (*ResvRTS*) packet sent by the sender in the next step. Thus, the reservation request process is continued only if all neighbors of the sender can receive the *ResvRTS* correctly, and can record the reservation.

3. Reservation request / Collision report: At the sender side, this step allows the sender to send its reservation request. It is executed by the sender only if no collision occurred during the 1st mini-slot at its neighbors, and if the sender receives correctly the *CTS* from the intended receiver. Thus, when the sender receives correctly the *CTS*, it sends a *ResvRTS* packet in the 3rd control mini-slot. Beside source and destination addresses, the *ResvRTS* includes the list of available slots at the sender, and the number of requested slots, and the class of service (real-time or best-effort). The list of available slots specifies the list of slots of the Data sub-frame available for transmission from the viewpoint of the sender. The number of requested slots specifies the requested bandwidth. At the receiver side, this step is used by the receiver neighbors to notify the receiver of collision during the 2nd control mini-slot. Thus, any neighbor of the receiver which senses collision during the 2nd mini-slot sends a Collision Report during the 3rd mini-slot in order to jam any reservation request (*ResvRTS*) packet destined to its neighbors. The aim behind this second *CR* is to ensure for the receiver that there is no two-hop neighbor which can transmit a reservation response (*ResvCTS*) in the 4th mini-slot at the same time, and the reservation is recorded by all neighbors of the receiver. When the receiver receives the *CR* or senses collision during the 3rd mini-slot, it remains silent during the 4th mini-slot and the reservation process is stopped. At the non reception of *ResvCTS* in the 4th mini-slot, the sender concludes that the reservation has failed and retries the reservation process in another *CRS*.

At the end of the third step, the reservation process is considered successful, and can be continued if no collision or *CR* is detected neither in the 2nd nor in the 3rd control mini-slot.

4. Reservation acceptation: If no collision occurred during the 2nd mini-slot at the receiver neighbors, the receiver will correctly receive the *ResvRTS* in the 3rd mini-slot. The receiver checks its *SST* and replies with a *ResvCTS* in the 4th mini-slot if there are common free slots between available slots list specified in the *ResvRTS* and its local reception available slots. The *ResvCTS* specifies the set of slots which will be reserved with the sender. Each node that receives the *ResvCTS* updates its *SST* and the set of available slots for transmission. This update prevents the node from reserving the specified slots for transmission.

5. *Reservation confirmation:* When the sender receives the *ResvCTS*, it replies with a *ResvConfirm* during the 5th mini-slot. The *ResvConfirm* indicates the set of slots which are reserved for transmission by the sender. Neighbors of the sender that hear the *ResvConfirm* update their *SST*, and do not accept reservation requests for the reserved slots.

Reservation steps are summarized in Figure 3-7.

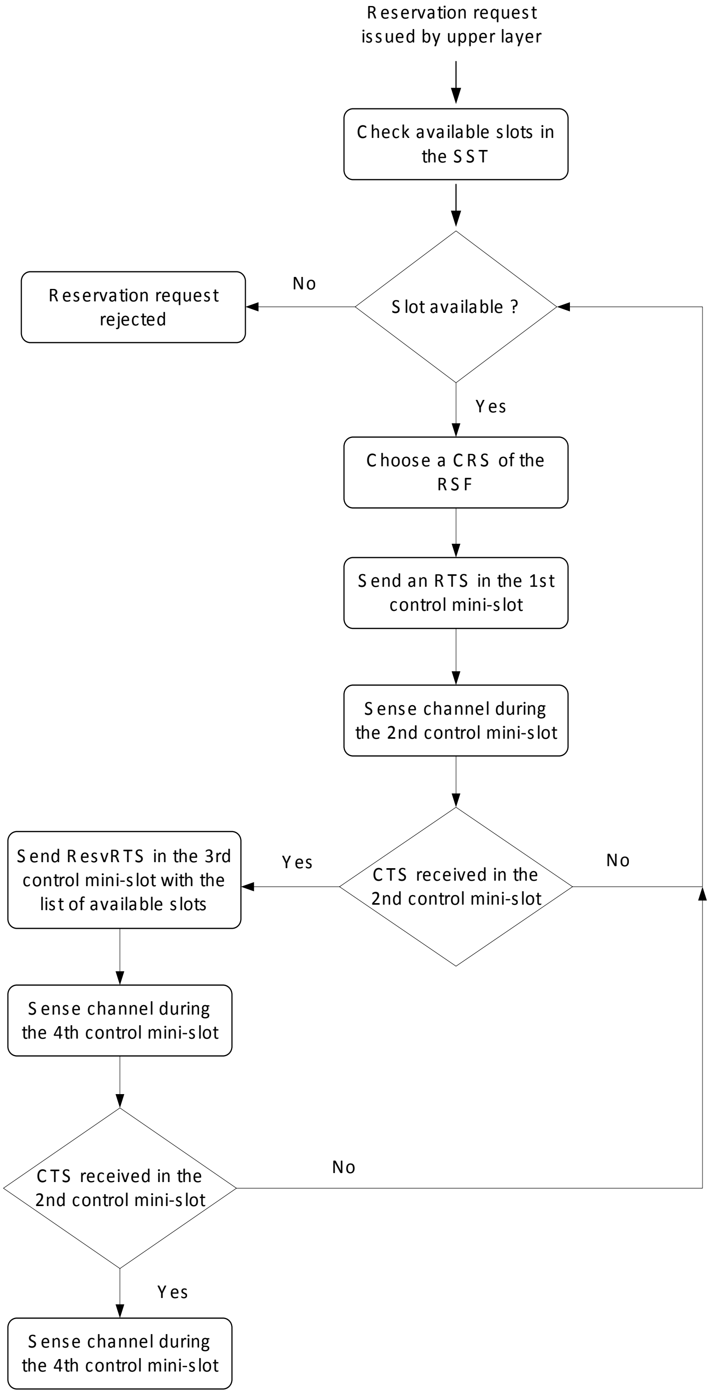

Figure 3-7. Reservation diagram of our protocol.

To illustrate this reservation steps, we consider scenarios of Figure 3-8. In the 1st scenario, only node B wants to reserve a slot with A. Since no collision is detected by neighboring nodes, the reservation is successfully established, and recorded by nodes C and F. In the second scenario, nodes B and D want to establish reservations with A and E respectively. If both B and D send their requests at the same time, then node C hearing a collision during the 1st mini-slot sends a *CR* in the 2nd mini-slot, and both nodes retry their reservation requests in another *CRS*. In the third scenario, nodes A and E want to reserve a slot with B and D respectively. If both of A and E send their requests on the same *CRS*, then nodes B and D reply with *CTS* simultaneously in the 2nd mini-slot. Node C hearing a collision during the 2nd mini-slot sends a CR in the 3rd mini-slot, and B and D will not receive the *ResvRTS* sent by A and E in the 3rd mini-slot. Both A and E defer their reservation requests.

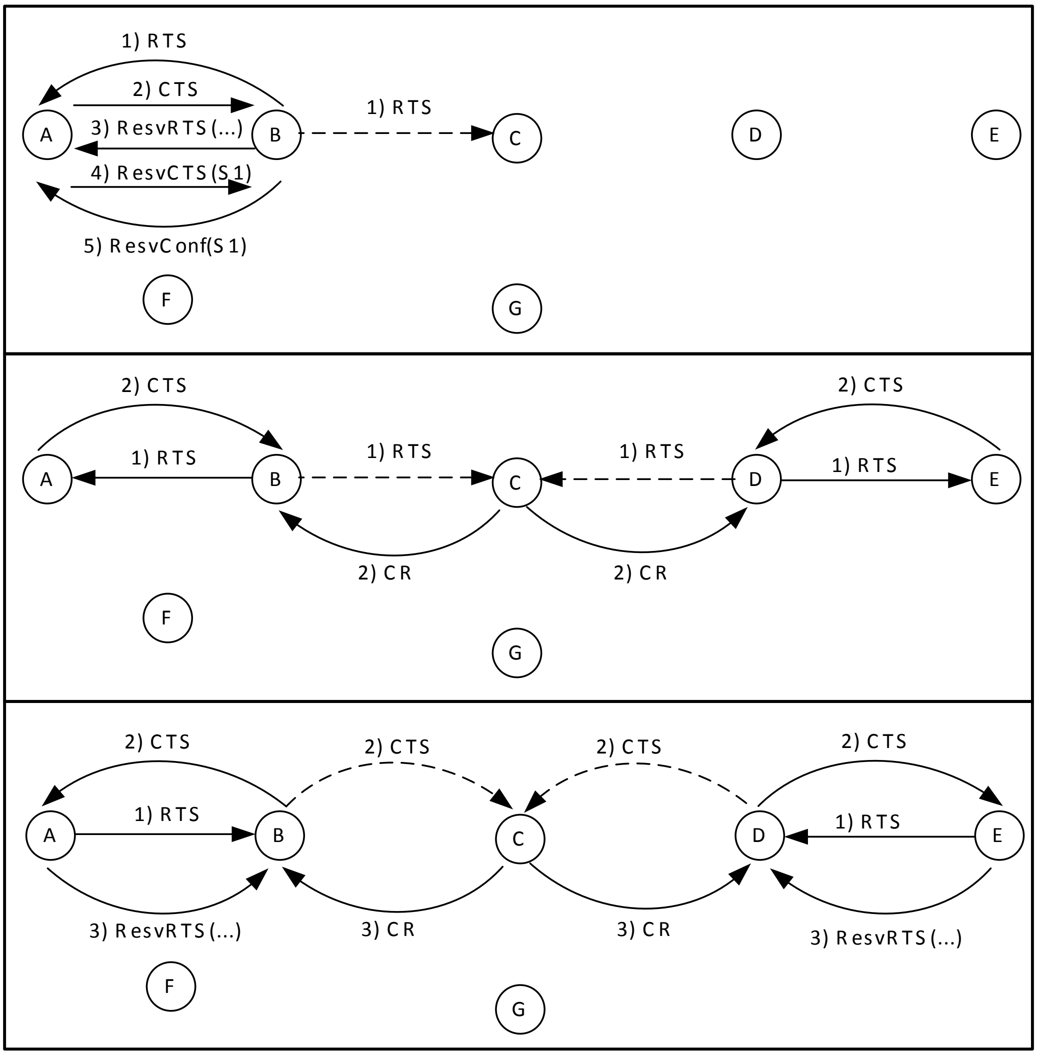

Figure 3-8. Reservation success and reservation failure scenarios during a *CRS*. The number before each control packet denotes the reservation step.

Through this handshaking scheme, any idle node in the network is able to notify the non-reception of reservation packets (reservation acceptation or reservation confirmation) in case of collision. Thus, the reservation scheme guarantees that once a reservation is accepted and confirmed, a single sender is allowed to send data packets free of collisions to a given receiver with high probability and with low risk of reservation loss.

Compared to other reservation schemes, that suffer the reservation breakage even in a low mobility network, this handshake scheme reduces scenarios where reservations are lost, and consequently reduces the probability of packet collision.

3.3.6 Frame formats

Packets received from upper layers are encapsulated into MAC sub-layer frames. Figure 3-9 shows the general MAC frame (noted MPDU) format. The frame control field includes information such as the frame type, and the "More Fragment" field. The "source MAC @" and "Dest MAC @" fields denote the MAC addresses of the source and the destination. The CRC is a 32-bits Cyclic Redundancy Check (CRC) code. We keep these fields as defined in the IEEE 802.11 standard description.

Acknowledgement frames are MAC frames generated by the MAC sub-layer in order to acknowledge the successful reception of data frames. Figure 3-10 shows the MAC sub-layer acknowledgement frame format.

Reservation Control packets format are shown in Figure 3-11.

MAC frames are encapsulated in PHY layer frames. The format of PHY layer frames is shown in Figure 3-12. The preamble and PHY header fields are physical layer dependent. The preamble consists of a sequence of bits used by the physical layer for synchronization, and as a start frame delimiter. The PHY header field includes information (such as the number of bytes in the frame, the transmission rate, and the CRC), used by the physical layer to decode the frame. The design of these fields is out of the scope of our work. Thus, we consider the values defined in the IEEE 802.11b physical layer description, i.e., 56 bits for the preamble, and 48 bits for the PHY header.

Octets:	2	6	6		4
	Frame control	Source MAC @	Dest MAC @	Frame body	CRC

Figure 3-9. MAC layer Data frame format.

Octets:	2	6	6	4
	Frame Control	Source MAC @	Dest MAC @	CRC

Figure 3-10. MAC layer Acknowledgement frame format.

2 bytes	6	6	4	1	4
Frame control	Source MAC @	Dest MAC @	List of available slots	Nbr requested slots	FEC

A) ResvRTS packet

2 bytes	6	6	4	4
Frame control	Source MAC @	Dest MAC @	List of reserved slots	FEC

B) ResvCTS packet

2 bytes	6	6	4	4
Frame control	Source MAC @	Dest MAC @	List of reserved slots	FEC

C) ResvConfirm packet

2 bytes	6	6	4	4
Frame control	Source MAC @	Dest MAC @	Released slots	FEC

C) ResvRelease packet

Figure 3-11. Reservation control packets format.

Preamble 144 bits	PHY Header 48 bits	MPDU

Figure 3-12. Physical layer frame format.

3.3.7 Reservation deadlock resolution

As illustrated in section 3.2, two neighbor nodes may use simultaneously the same *CRS* to transmit their reservation requests. The two nodes are not able to hear the reservation packets sent by each other, and cannot record reservation. We called this scenario the deadlock problem.

With the handshake scheme described in section 3.3.5, the deadlock problem can be easily resolved if the two nodes transmitting control packets simultaneously have at least one common neighbor. Such a neighbor participates in resolving the deadlock situation through sending a Collision Report, thus, notifying the collision of reservation packets.

To illustrate this situation, consider the scenario of Figure 3-13 where nodes A, C, and E are all one-hop neighbors. Thus, if nodes A and C chose the same *CRS* for slot reservation, the deadlock happens because A and C cannot receive the *RTS* packet transmitted by each other. The reservation deadlock in this situation is resolved during the handshake. Indeed, node E will hear collision of *RTS* packets during the 1st control mini-slot, and sends a *Collision Report* in the 2nd control mini-slot of the same *CRS*. Consequently, nodes A and C will cancel the current handshake and retry the reservation later.

The deadlock problem is more challenging if the two nodes A and C do not have a common neighbor. To illustrate what happens in this situation with our protocol, consider the scenario of Figure 3-14. In this scenario, both of A and C establish a reservation with B and D respectively, but nodes A and C cannot receive the reservation confirmation sent by each other as they send their confirmations simultaneously. In addition, there is no common neighbor that hears and signals the collision. Consequently, both of A and C remain unaware of the reservation made by each other, and one of them may later break the reservation of the other, causing collision.

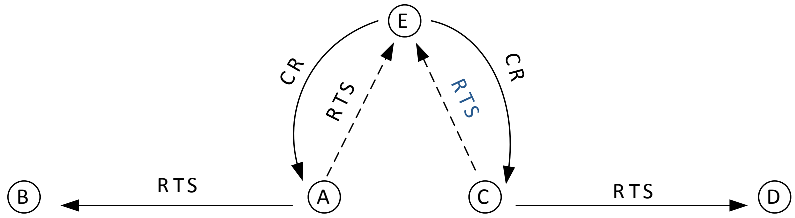

Figure 3-13. Reservation deadlock avoided during the reservation handshake.

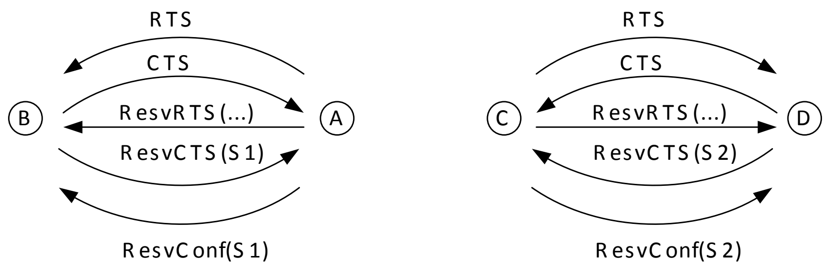

Figure 3-14. Reservation deadlock cannot be resolved during the reservation handshake.

Basically, regarding the characteristics of the wireless transceivers, there is no solution for a node to sense the channel while it is transmitting a packet. Consequently, our protocol cannot resolve the reservation deadlock situations during the reservation establishment phase if the sender and the receiver have no common neighbors. Alternatively, we propose a reactive solution that consists in resolving the reservation conflict caused by the reservation deadlock as soon as the deadlock causes a collision during reserved slots.

The deadlock situation is detected when a node that accepted a reservation request senses collision during its reception reserved slot. In this situation, the receiver understands that its reservation is being lost, and the slot must be released with the corresponding sender. Furthermore, the receiver should inform its neighbors that the slot is no more considered reserved by the receiver, and consequently can be reserved for transmission by these neighbors. Indeed, the conflict of reservation is resolved by making the receiver sending a *ResvRelease* packet to indicate to the sender that the reservation is lost. The *ResvRelease* specifies slots on which the collision occurred. When the sender receives the *ResvRelease*, it stops sending packets on the reserved slot, and reserves another slot with the intended receiver. When the receiver neighbors receive the *ResvRelease*, they update their slots state tables, and the slot is considered available for transmission.

In order to ensure that the sender and all neighbors of the receiver receive the *ResvRelease* collision-free, the *ResvRelease* is sent in the same way as *ResvRTS* is transmitted, during the first control mini-slot of a *CRS*. Hence, any node that hears collision of the *ResvRelease* sends a *CR*, and the sender retransmits its *ResvRelease* until successfully received by all neighbors.

3.4 Correctness proof

In this section, we assess the effectiveness of the proposed protocol in maintaining consistency of reservations in environment without mobility.

Property 3-1. The protocol resolves the scenario 1 of unheard reservation control packets.

Scenario 1 (cf. Figure 3-1) happens when an idle node does not receive the reservation acceptations (*ResvCTS*) sent by its neighbors due to their collision.

We can prove that this scenario cannot happen with our reservation scheme, if we prove that two nodes with a common idle neighbor cannot transmit a *ResvCTS* simultaneously.

Proof. In Figure 3-1, with our protocol, node E3 does not receive the *ResvCTS* packets sent by R1 and R2 if they collide at node E3. R1 and R2 send *ResvCTS* in the 4th mini-slot only if they have sent *CTS* in the 2nd mini-slot, and received a *ResvRTS* in the 3rd mini-slot. However, R1 and R2 cannot receive *ResvRTS* in the 3rd control mini-slot (and consequently they cannot send *ResvCTS* in the 4th mini-slot) because as they send *CTS* in the 2nd mini-slot simultaneously, collision occurs at their common neighbor E3 which sends a CR in the 3rd mini-slot, preventing R1 and R2 from receiving the *ResvRTS*. Consequently, the scenario 1 of the unheard control packets cannot occur.

Property 3-2. The protocol resolves the scenario 2 of unheard reservation control packets.

This scenario occurs when an idle node cannot record a transmission reserved slot, because two (or more) of its one-hop neighbors send their reservation confirmation at the same time. In order to prove that this scenario cannot happen with our reservation scheme, we need to prove that two nodes that have a common idle neighbor cannot transmit a *ResvConfirm* simultaneously.

Proof. Let us consider Figure 3-2 that clearly illustrates this scenario. With our scheme, nodes E1 and E2 send *ResvConfirm* only if they received a *ResvCTS* in the 4th mini-slot, sent *ResvRTS* in the 3rd mini-slot, and received *CTS* in the 2nd mini-slot. However, E1 and E2 cannot receive *CTS* from their intended receivers because as they send *RTS* simultaneously in the 1st mini-slot, collision occurs at their common neighbor R3 which sends a CR in the 2nd mini-slot, preventing them from sending their *ResvRTS*. Consequently, E1 and E2 cannot send *ResvConfirm*, and the scenario 2 of the unheard control packets cannot occur.

Property 3-3. In a network without mobility, scenarios 1 and 2 cannot be the cause of reservation loss and the unique cause of collision during reserved slots is reservation deadlock.

Suppose that a node E1 reserved slot S1 with node R1. Collision occurs on slot S1 at R1 if there exists another node E2, neighbor of R1, which reserves slot S1 for transmission. This may occur if node E2 has not heard the *ResvCTS* sent by R1 because E2 was transmitting when R1 sent the *ResvCTS* (which represents the reservation deadlock scenario), or if the *ResvCTS* sent by R1 collided with another *ResvCTS* sent by another neighbor of E2 (which represents the scenarios 1 and 2). However, the latter case cannot occur because the CR sent by E2 in the 3rd mini-slot prevented R1 and R3 from receiving the *ResvRTS* and the reservation could not be established. Consequently, scenarios 1 and 2 as a cause of collision is eliminated, and the unique cause of collision in a network without mobility are reservation deadlock scenarios.

Property 3-4. The protocol allows nodes that have lost their reservations due to reservation deadlock to detect the reservation loss, release their lost reservation, and establish new reservations.

Proof. The proposed protocol does not avoid the scenarios of reservation deadlock, but it defines a reactive mechanism which permits nodes that were not able to hear control packets sent by their neighbors to detect the reservation loss.

Consider the reservation deadlock scenario of Figure 3-3 where node E1 has not heard the reservation confirmation sent by E2 for slot S2. When E1 accepts the reservation request of E4 on slot S2, collision occurs on slot S2 at E1. As a result, E1 and E4 lose their reservation. E1 sends a *ResvRelease* to release slot S2 with E4 and with its neighbors. Once the reservation is released, nodes E1 and E4 retry to establish a reservation on another slot.

Similarly, in reservation deadlock of scenario of Figure 3-4, node R1 has not heard the *ResvCTS* of R2 for slot S2. Later, R1 establishes a reservation with R4 on slot S2. When R1 transmits its packets, collision occurs at R2 on slot S2. R2 detects the collision and releases slot S2 with E2. E2 releases the reservation and reserves another slot.

3.5 Performance evaluation

In this section, we are interested in analyzing through simulation, whether our protocol provides better performance than existing protocols. Particularly, we compare the performance of our protocol with one synchronous protocol, i.e, R-CSMA, and an asynchronous protocol, namely RTMAC. The objective of this evaluation is to observe the behavior of the protocol and determine its ability to provide low reservation breakage rate, low collision. In addition, we seek to evaluate the overhead of our solution compared to the overhead engendered by the other solutions.

We don’t consider any application area. The following chapters contain more material in order to take into consideration a specific application field. We consider a wireless channel of 2 Mb/s, and a static Ad-hoc network composed of CBR traffic sources. Each CBR session generates a 100 bytes packet every 100 ms. The super-frame length is set to 100 ms, and the slot length to the transmission time of one CBR packet with the different layers overhead. Each CBR session reserves one slot to accommodate its traffic and stays 30 s in the system after which it releases its reserved slot.

Table 3-1. Simulation parameters.

Parameter	Value
Channel bit rate (Mbps)	2
UDP+IP header (bytes)	8+20
MAC header (bytes)	18
PHY layer overhead (PLCP header + preamble) (bits)	48+56
Data slot payload size (bytes)	100
Data Slot length (L_{slot}) (ms)	0.626
Guard time between slots (μs)	2
Super-frame length (SFL) (ms)	100
Number of CRS in Reservation Sub-Frame	20
number of data slots per super-frame	55
RTS length (bytes)	18
CTS length (bytes)	18
ResvRTS length (bytes)	23
ResvCTS length (bytes)	22
ResvConfirm length (bytes)	22
RTS mini-slot length (ms)	0.124
CTS mini-slot length (ms)	0.124
ResvRTS-mini-slot length (ms)	0.144
ResvCTS-minislot length (ms)	0.140
ResvConfirm length (ms)	0.140
CRS length (ms)	0.7
Simulation time (s)	1000s

3.5.1 Simulation results

The occurrence of reservation conflicts is expected to increase with the increase of contending nodes for slot reservation. In the rest of this section we analyze the impact of traffic load and the reservation request rate on the reservation failure, collision rate, and control overhead.

We suppose that CBR sessions arrive into the network and start sending their reservation requests following a call arrival rate (i.e., number of calls per second).

Figure 3-15 the reservation breakage frequency per connection versus the increase of the call arrival rate obtained by R-CSMA, RTMAC, and our protocol. It shows that reservation breakage occurs less frequently with our protocol. At an arrival rate of 12 connection/s, a connection losses its reservation every 8 seconds with our protocol, while with R-CSMA a connection losses its reservation every 4 seconds. With RTMAC reservation breakage occurs every 3 seconds. Thus, our protocol is less affected by the increase of traffic load compared to R-CSMA and RTMAC.

Figure 3-16 shows the collision rate versus the increase of the call arrival rate. As one may notice, our protocol performs lower collision rate compared to R-CSMA and RTMAC, especially at high call arrival rate. In particular, the collision rate of our protocol is as short as 4 collisions per second when the traffic is maximal in comparison to 10 collisions per second with R-CSMA and 13 collisions per second with RTMAC.

Figure 3-17 shows the control overhead engendered by slot reservation versus the increase of the call arrival rate. Our protocol generates less control traffic than R-CSMA and RTMAC despite it uses a five-way handshake reservation. The low overhead of our protocol in comparison with the other protocols is explained by its low reservation breakage rate. As reservation breakage with our protocol occurs less frequently, nodes are not required to retry the reservation request process so frequently during the life time of the connection, and consequently they generate less traffic. Contrarily, as R-CSMA and RT-MAC have higher reservation failure rate; nodes are required to send their reservation requests each time a reservation is broken, explaining the high control traffic generated by these two protocols.

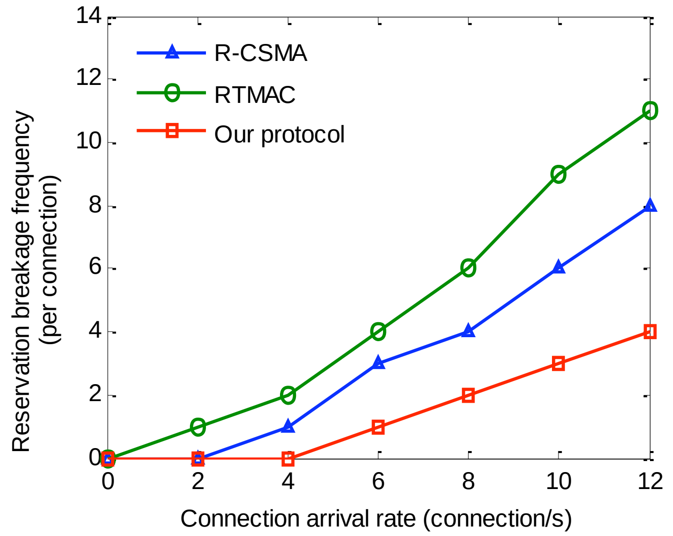

Figure 3-15. Reservation breakage frequency vs. the increase of connection arrival rate.

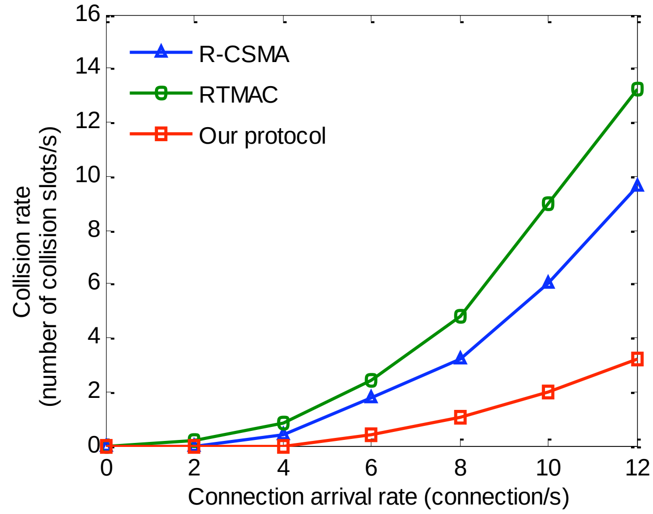

Figure 3-16. Collision rate vs. the increase of connection arrival rate.

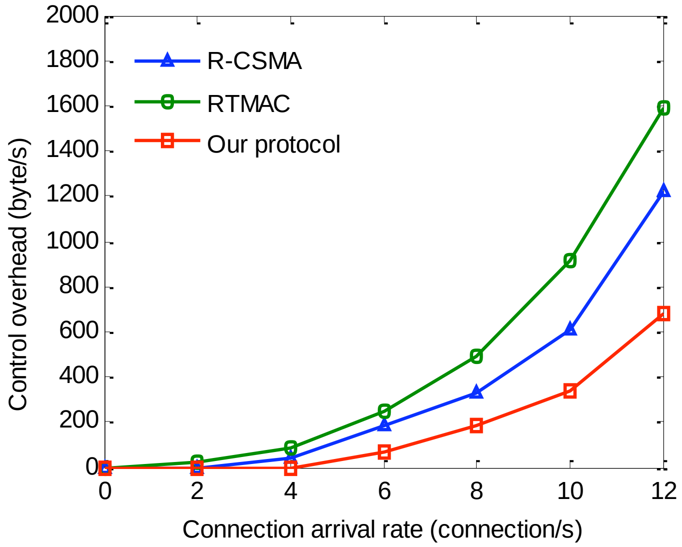

Figure 3-17. Control overhead vs. the increase of connection arrival rate.

3.6 Discussion

Consistency of reservation is an important feature that impacts the performance of any reservation protocol. Reservation inconsistency refers to the scenarios where some nodes of the network violate the collision-free reservation conditions, and where conflicts of reservations may occur.

In this chapter, we identified the major causes of reservation conflicts, and the major issues against achieving collision-free transmission in reservation MAC protocols. We illustrated that these conflicts are mainly due to the different scenarios in which nodes are not able to hear reservation control packets sent by their neighbors.

The reservation protocol that we propose aims at resolving these conflicts, and establishing consistent reservations in order to reduce the reservation breakage. The main idea of the proposed protocol is to provide better coordination between nodes. The protocol ensures that each reservation control packet transmitted by a node to establish a reservation is successfully received by all its neighbors.

The protocol presented in this chapter represents a basic reservation scheme that can be used to provide QoS for a variety of multimedia and real-time applications. The next step of our work consists in analyzing the performance of the protocol in providing QoS for voice traffic. The first step (Chapter 4), consists in exposing the basic components of voice communication over wireless Ad-hoc networks architecture, and the characteristics and QoS requirements of voice traffic. This step will allow us to understand better voice traffic requirements, and choose the protocol parameters (i.e., the data-slot length, the length of the reservation sub-frame in terms of the number of *CRS*, and the permission probability of nodes to transmit their reservation requests) that give the best results for voice traffic (cf. Chapter 5). Furthermore, we propose mechanisms for voice/data integration over the proposed reservation scheme. Indeed, this step aims at determining how both voice and data traffics share and access the data sub-frame so that the performance of both types of traffic are improved. The following step (Chapter 6) consists in evaluating the performance of the protocol through both simulation and analytical models.

Chapter 4: Voice Communication over MANETs

4.1 Introduction

Among the services which are expected to emerge over MANETs, we have voice communication service. Thus, MANETs have the ability to provide cost effective voice service, allowing the integration of voice and data communication over the same network.

Real-life applications where voice/data communication over MANETs is useful are environments where voice communication is vital and where the deployment of wired infrastructure is unfeasible such as firefighting, life-saving operations, and tactical networks. However, Voice over MANETs poses significant challenges since the bandwidth in MANETs is much limited than that in wired counterparts. The physical and MAC layers designed for MANETs are still unable to provide efficient utilization of the limited bandwidth. For example, existing standards (such as IEEE 802.11 [IEE 05]) cannot answer the QoS requirements of voice since they were not originally designed to support delay-sensitive applications. Thus, QoS provisioning for voice traffic support over MANETs is still a challenge, and an efficient solution for voice/data integration in MANETs is still an open research field.

Before designing a solution for voice/data communication over Wireless Ad-Hoc Networks, it is necessary to understand the characteristics of voice traffic, its QoS metrics, and the major issues in QoS provisioning for voice traffic in wireless networks. In this chapter, we provide a description of the Voice/data communication over wireless ad-hoc networks architecture. Firstly, we present the components composing this architecture. Then, we present the major concepts of QoS for voice traffic, their definition, their metrics, and the principal issues in providing QoS for voice traffic over MANETs. We end the chapter by an overview of the impact of the MAC sub-layer on the performance offered by the infrastructure to voice traffic.

4.2 Network Architecture for Voice/Data Communication over Wireless Ad-hoc Networks

Figure 4-1 depicts a generic architecture for voice/data transmission over wireless ad-hoc networks, which consists of the voice coding/decoding sub-systems, communication sub-system, and the wireless network. Among these sub-systems, the communication sub-system is a key component of this architecture. Some other components such as loss/error concealment, playout buffer for speech smoothing, may also be included in the system to enhance the functionality and performance of the infrastructure. However, these functionalities are not considered in our work.

A voice communication within this architecture is accomplished through three phases: connection establishment, data transfer, and connection tear-down. In the connection establishment phase, the source and the destination perform "handshaking" to establish a logical connection between the two peer entities. Once the connection establishment phase is complete, the voice traffic transfer phase begins. The voice coding sub-system encodes the incoming voice analog signal in real-time using a voice codec. The corresponding encoded voice stream is then packetized and sent to the communication sub-system which controls its transmission over the wireless network. At the receiving side, the reverse process of de-packetization and decoding is performed to reproduce the original voice analog signal. Finally, at the end of the voice conversation, the virtual connection is "torn down".

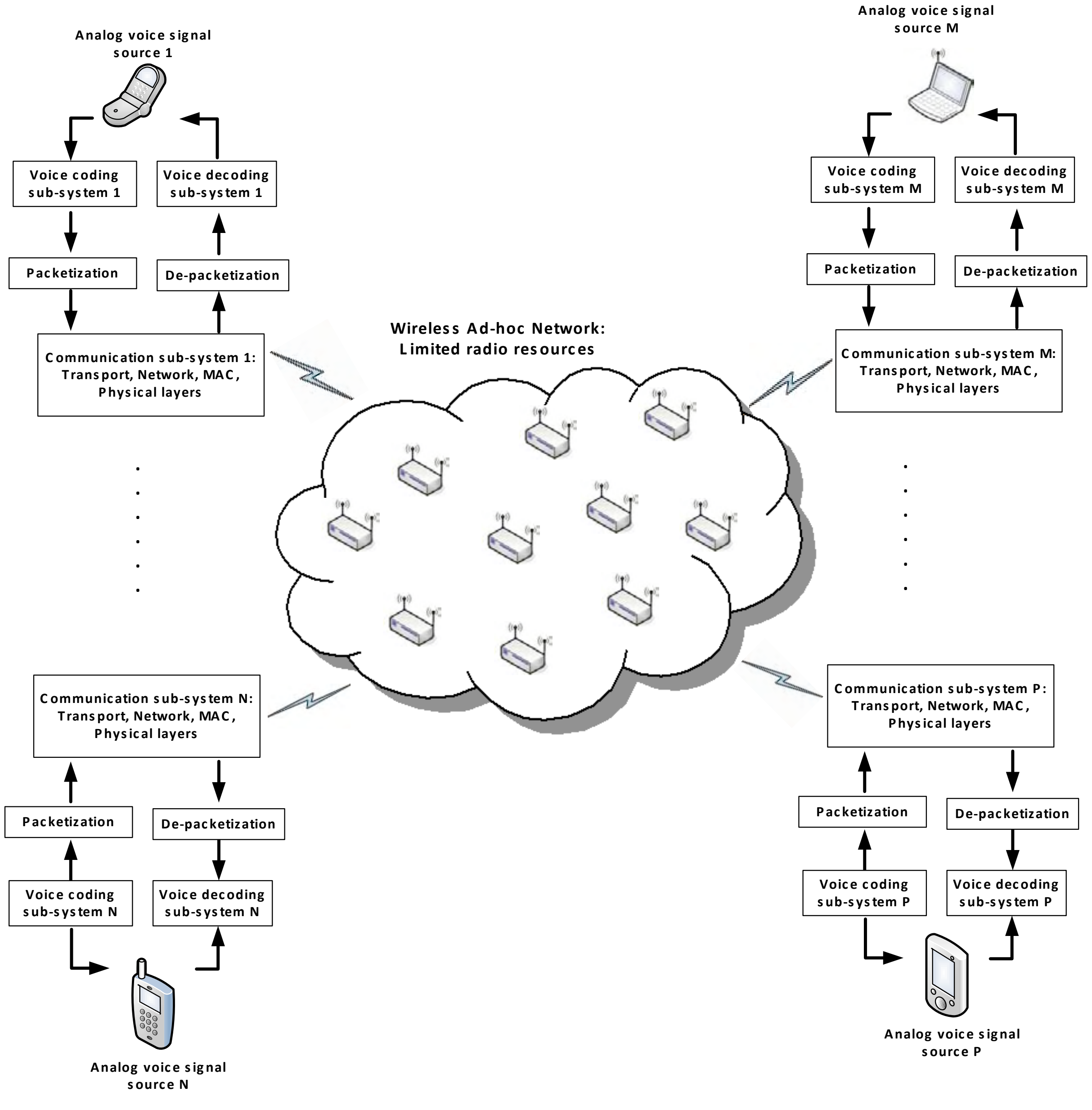

Figure 4-1. Generic architecture for voice/data communication over wireless ad-hoc networks.

4.2.1 Voice coding sub-systems

Voices and sounds heard by human ear are analog signals. Their transmission on any digital medium requires their conversion to digital format, suitable for transmission over digital networks (MANETs, LAN, internet ... etc). This voice conversion process from the analogical format to the digital one is called encoding. The task of encoding is performed by voice codecs.

In some network technologies where bandwidth is limited (such as wireless networks) it is highly desirable to reduce the bandwidth required for the transmission of the digitized speech signal. To reduce the required bandwidth, some voice codecs use compression and redundancy removing algorithms to reduce the bit rate of the encoded voice stream. After coding and compression, the codec generates a fixed bit rate voice stream.

There are two types of speech coding techniques: waveform coding and linear predictive coding (LPC) [KUN 05].

4.2.1.1 Waveform coding

The waveform coding is a fairly straightforward technique in which the analog signal is sampled at discrete time points and each sample is represented by a number of bits. The waveform coding takes its mathematical basis from the Nyquist theorem. The Nyquist theorem suggests that if an analog signal is band-limited by *B* KHz, then the analog signal can be sampled at twice the bandwidth *B* and the original analog signal can be reconstructed from the discrete list of samples. The Pulse Code Modulation (PCM) and the Adaptive Differential Pulse Code Modulation (ADPCM) algorithms are two examples of waveform coding [KUN 05].

The Pulse Code Modulation (PCM): Originally introduced by AT&T Bell laboratories, the PCM is the first speech coding algorithm and is the most widely used. Following the Nyquist theorem, as normal human speech signal is band-limited by 4 KHz, the speech signal can be sampled at the rate of 8 KHz. Thus, in the PCM algorithm, analog speech signal is sampled at the rate of 8 KHz yielding 8 000 samples per second, with each discrete sample is coded into eight bits. This results in a 64 Kbps digital bit stream representing the source speech signal:

$$(8000\, samples/\mathrm{sec}) \times 8\, bits/sample = 64\, Kbps \tag{4-1}$$

The frame length of the PCM is of one sample, i.e., eight bits, and the frame duration is 0.125 ms. The PCM represents the basis of several enhanced voice codecs.

The Adaptive Differential Pulse Code Modulation (ADPCM): The ADPCM was introduced to reduce the bit rate of the PCM in order to save bandwidth. ADPCM uses a sample prediction algorithm. First, each sample instant t_N the algorithm uses the *(P-1)* previous samples to predict the N^{th} sample. Then, the difference between the actual sample and its predicted value is calculated:

$$S_N^* = \sum_{i=N=P}^{N=1} \alpha_i S_i \tag{4-2}$$

$$\Delta_N = S_N^* - S_N \tag{4-3}$$

Where

S_i is the amplitude of the i^{th} sample;

S^*_N is the predicted amplitude of the N^{th} sample;

α_i Coefficients of the linear prediction filter;

Δ_N The difference between the amplitude of the N^{th} sample and its predicted amplitude;

P The length of the prediction filter;

With the PCM, for each sample, the amplitude (S_N) is encoded into eight bits. ADPCM uses a differential coding technique to reduce the number of bits required for the encoding of samples. Instead of coding the whole amplitude of a sample, the ADPCM considers only the difference between the sample amplitude (S_N) and its predicted amplitude (S^*_N), i.e., Δ_N.

Since the adjacent speech samples tend to have close amplitudes, the difference between the amplitude of a sample and its predicted amplitude tends to be much smaller than the absolute amplitude of the sample. Thus, since the number of bits required for encoding the difference between the predicted amplitude and the actual sample amplitude is smaller than those required for encoding the sample amplitude, the differential coding technique used in the ADPCM can reduce the bit rate compared to the PCM rate.

Like the PCM, the ADPCM samples speech at 8,000 samples/sec rate. However, in the ADPCM, each sample is coded into four bits instead of eight bits yielding 32 Kbps. This bandwidth reduction is obtained at the expense of an increase of codec complexity due to the prediction algorithm.

4.2.1.2 Linear Predictive Coding (LPC)

Linear Predictive Coding (LPC) [KUN 05] is one of the powerful speech analysis techniques, and one of the most useful methods for encoding good quality speech at a low bit rate. The LPC uses an approach based on accurate estimation of speech parameters. Instead of sending the samples representing a speech signal and reconstructing the original speech from the received samples, the LPC learns the individual speaker vocal tract characteristics and estimates the model parameters in real-time based on the analysis of speech samples. The estimated parameters are sent to the receiver for decoding. The decoder in the receiver uses the received parameters to reconstruct the speech model and synthesize the speaker voice. With the LPC, drastically lower bit rates have been achieved.

The Code-Excited Linear Predictive (CELP) coding technique is a class of LPC coding. The CELP family of codec includes the Conjugate-Structure Algebraic Code-Excited Linear Predictive (CS-ACELP) standardized by the ITU-T G.729 [ITU 07] and the Low-Delay Code-Excited Linear Prediction LD-CELP standardized by the ITU-T G.728 [ITU 92].

4.2.1.3 Voice coding standards

Many codecs have been standardized by the ITU-T in its G-series recommendations and by other organizations. Each codec is designed for a specific use, and has its own properties such as the offered bit rate, the sampling rate, and the frame duration. The criterion of performance evaluation of voice codecs are the sound quality, and the computational complexity of the algorithm used for coding/decoding.

The sound quality is measured by the *MOS* (*Mean Opinion Score*) [KUN 05]. The MOS ranges from 1 to 5, with 1 corresponding to the worst quality and 5 corresponding to an excellent quality. The complexity of a voice codec is measured in *MIPS* (*Million of Instruction Per Second*).

Table 4-1. MOS (Mean Opinion Score).

MOS score	Verbal rating
5	Excellent
4	Good
3	Fair
2	Poor
1	Unsatisfactory

The most representative codecs used in wireless communication are:

• ***G.711:*** Also known as a-law/μ-law, the G.711 codec [ITU 93] is one of the oldest voice encoding schemes used in VoIP communications. Standardized by the ITU-T in 1988, it uses the Pulse Code Modulation (PCM) that samples the source speech signal at a rate of 8 KHz, with each sample is coded into 8 bits. This results in 64 Kbps digital bit stream representing the source speech signal. This codec works best in networks where a lot of bandwidth is available. Its benefits include simple implementation which does not need much CPU power, and a very good perceived audio quality with a MOS of 4.4.

• ***G.722:*** G.722 [ITU 88] is the 64 Kbps ITU-T standard for wideband applications. It describes a 50 to 7 000 Hz audio coding system which may be used for a variety of higher quality speech applications. The coding system uses Sub-Band Adaptive Differential Pulse Code Modulation (SB-ADPCM) algorithm with a bit rate of 64 Kbps. The SB-ADPCM technique splits the 8 KHz frequency band into two sub-bands (lower sub-band (0-4000 Hz) and higher sub-band (4000-8000 Hz)) and the signals in each sub-band are sampled at a rate of 16 KHz and encoded using ADPCM (Adaptive Differential Pulse Code Modulation). Prior to encoding, each sub-band signal is down-sampled by a factor of two.

The higher band resolution is fixed at 2 bits/sample, while the number of bits allocated to the lower band can be adjusted in order to achieve different coding rates. Hence, the encoder can operate in three different modes: 64, 56, and 48 Kbps corresponding to 6, 5, or 4 bits/sample in the lower band respectively.

• ***G.723.1:*** [ITU 06] specifies a coder that can be used for compressing the speech signal at a very low bit rate using a limited amount of complexity. The coder operates on speech frames of 240 PCM samples, i.e., 30 ms frame length. It encodes every frame into 10 or 12 16-bit code-words yielding 5.3 Kbps or 6.3 Kbps rates respectively. Thus, audio signal is encoded in 30 ms frames that can be of 24 bytes size for 6.3 Kbps rate, or 20 bytes for 5.3 Kbps rate. The 6.3 Kbps rate is achieved using the MPC-MLQ (Multipulse LPC with Maximum Likelihood Quantization) codebook search algorithm, while the 5.3 Kbps rate is obtained using the ACELP algorithm.

The 6.3 Kbps rate has greater quality compared to the 5.3 Kbps rate. However, the lower bit rate gives good quality and provides system designers with more flexibility. It is possible to switch between the two rates at any 30 ms frame boundary. The complexity of the algorithm is rated at 25 MIPS.

• ***G.726:*** G.726 [ITU 90] describes the algorithm recommended for conversion of a 64 Kbps a-law or µ-law PCM stream to and from a 40, 32, 24, or 16 Kbps bit rate stream. The conversion is applied to the PCM stream using the ADPCM transcoding algorithm. The PCM samples stream is transformed into a stream of codewords with a one-to-one correspondence to the samples in the PCM stream. Codewords can be coded into 5, 4, 3 or 2 bits, with the sign of the difference is coded on one bit and the amplitude on 1, 2, 3 or 4 bits, resulting in bit rates of 16, 24, 32 or 40 Kbps respectively.

Prior to 1990, G.721 described the 32 Kbps ADPCM encoding, and G.723 described the 40, 32, and 16 Kbps encodings. Thus, G.726 with 32 Kbps rate designates the same algorithm as G.721 in RFC 3551 [SC 03].

• ***G.728:*** The G.728 codec [ITU 92] encodes speech at 16 Kbps rate using Low-Delay CELP (LD-CELP) algorithm which uses a prediction scheme that consists in coding the difference between the real sample value and its estimation based on the previous sample value. G.728 encoder translates 5 consecutive audio samples into a 10-bit codebook index, resulting in a bit rate of 16 Kbps for audio sampled at 8 000 samples per second. The group of five consecutive samples is called a vector.

• ***G.729:*** The G.729 codec [ITU 07] also referred to as CS-ACELP (Conjugate Structure Algebraic Code Excited Linear Prediction), encodes speech at a low rate of 8 Kbps stream to make more efficient usage of the network bandwidth. It uses the Conjugate-Structure-ACELP (CS-ACELP) algorithm. The coder operates on speech frames of 10 ms corresponding to 80 samples at a sampling rate of 8 000 samples per second. For every 10 ms frame, the speech signal is analyzed to extract the parameters of the CELP model (linear-prediction filter coefficients, adaptive and fixed-codebook indices and gains) [HAR 03]. These parameters are encoded and transmitted, and used by the decoder to reproduce the original speech. After computing the reconstructed speech, it is further enhanced by a postfilter.

• ***GSM Full Rate (GSM-FR):*** Also called GSM 06.10, *GSM-FR* codec [RWO 95] is the speech coding standard designed by the European Telecommunications Standards Institute (ETSI) for GSM digital cellular systems. The coding algorithm used is RPE-LTP (Regular Pulse Excitation - Long Term Prediction). The encoder takes its input as 13 bit uniform PCM signal. GSM 06.10 recommendation describes the mapping between input blocks of 160 speech PCM samples to encoded blocks of 260 bits and from encoded blocks of 260 bits to output blocks of 160 reconstructed speech samples. The sampling rate is 8000 sample/s leading to a bit rate for the encoded bit stream of 13 Kbps.

• ***GSM Enhanced Full Rate (GSM-EFR)*** *[RWO 95]****:*** Described in recommendation GSM 06.60, the GSM-EFR encoder describes the mapping between input blocks of 160 speech samples in 13-bit uniform PCM form to encoded blocks of 244 bits and vice versa. The sampling rate is 8 000 samples/s leading to a bit rate for the encoded bit stream of 12.2 Kbps. The mapping is performed through the ACELP algorithm. The coder operates on speech frames of 20 ms corresponding to 160 samples at the sampling frequency of 8000 sample/s. At each 160 speech samples, the input PCM samples are analyzed to extract the parameters of the CELP model.

- ***iLBC (internet Low Bitrate Codec):*** The iLBC [AND 04] is a free speech codec developed by Global IP Solutions (GIPS) for a variety of VoIP applications such as streaming audio, archival and messaging. It is meant to cover the narrow frequency range of 90-4000 Hz. The codec uses a Block-Independent LPC (BI-LPC) algorithm and has support for two frame lengths: 20 ms at 15.2 Kbps and 30 ms at 13.33 Kbps. The input of the encoder must be PCM speech signal sampled at 8 KHz with each sample encoded onto 16 bits (i.e., a bit rate of 128 Kbps). Thus, the iLBC encoder compresses the input speech signal to 10.4 % and 11.9 % in order to produce a 13.33 Kbps or 15.2 Kbps digital signals respectively. The iLBC codec is being used by many PC-to-Phone applications, such as Skype, Google Talk, Yahoo! Messenger, and MSN.

A summary of voice codecs and their characteristics are listed in Table 4-2. It should be noticed that the values of the MOS are based on experimental tests, and there may be minor differences from one test to another following the test conditions.

Table 4-2. Characteristics of different voice codecs.

Codec	Creator	Algorithm	Data rate (Kbps)	MOS	Complexity (MIPS)
G.711	ITU-T	PCM	64	4.1	0.34
G.722	ITU-T	ADPCM	48		
G.723.1 - 6.3Kbps	ITU-T	MP-MLQ	6.3	3.87	25
G.723.1 - 5.3 Kbps	ITU-T	ACELP	5.3	3.69	25
G.726 - 40 Kbps	ITU-T	ADPCM	40	3.85 to 4.5	14
G.726 - 32 Kbps			32		
G.726 - 24 Kbps			24		
G.726 - 16 Kbps			16		
G.728	ITU-T	LD-CELP	16	3.61	
G.729	ITU-T	CS-ACELP	8	4.0	10
GSM FR	ETSI Special Mobile Group	RPE-LTP	13	3.5	2.5
GSM EFR	ETSI Special Mobile Group	CELP	12.2	3.8	15.4
iLBC 20 ms	Global IP Solutions	LPC	15.2	3.9	
iLBC 30 ms	Global IP Solutions		13.33		

The choice of the codec is a compromise between the desired voice quality and the capacity of the network infrastructure to offer the required bandwidth. An important factor impacting this choice in wireless networks is the required bandwidth which impacts the number of simultaneous voice conversations. The G.711 codec has high bandwidth requirements and gives better quality, while the other codecs are interesting in term of bandwidth consumption at the expense of a reduction of the speech quality as shown in Table 4-2.

If bandwidth is not an issue then the traditional codec G.711 is the best choice to use. It is the only codec that has achieved an excellent grade of service. If bandwidth is an important issue, then a compromise need to be made between bandwidth requirements, voice quality, and additional delay resulting from the use of the codec. In practice, G.729 and G.723 are the most popular compression schemes used.

4.2.1.4 Voice Activity Detector (VAD)

Generally, during a voice conversation only one interlocutor is speaking, the other interlocutor is hearing and consequently silent. During silence periods, only the signal corresponding to the background noise is present. The signal corresponding to the noise can be differentiated from the speech signal through its low strength, and it is possible for voice codecs to detect silence periods and to not transmit anything during these periods. This process of silence periods detection is called Voice Activity Detection, and allows a bandwidth conservation of up to 60% of the nominal bandwidth [BAH 06].

4.2.2 Packetization/De-packetization

Once the voice speech digitized, the packetizer collects the voice frames or samples from the encoder and generates a voice packet each packetization interval. The packetization interval determines the number of samples included within a single packet. Voice packets are encapsulated in the Real-time Transport Protocol (RTP) which is an UDP-based protocol. Each voice packet is carried in an RTP packet (with a 12 bytes header), which is carried in UDP (with an 8 bytes header), which is carried in IP (with a 20 bytes header). At the reception side, received voice packets are de-packetized to extract the coded speech signal (cf. Figure 4-2).

The voice packetization process incurs an overhead due the RTP/UDP/IP/MAC layers headers. Given a codec and the packetization interval *(PI)*, the overhead is calculated as follows:

$$Overhead\ of\ voice\ packetization = \frac{Codec\ rate \times PI}{(Codec\ rate \times PI) + size(RTP\ header) + size(UDP\ header) + size(IP\ header) + size(MAC\ header)} \quad (4\text{-}4)$$

The packetization interval has an important effect on the voice quality and bandwidth utilization of the network. There is a compromise between low packetization overhead and high voice quality. On one hand, using shorter packetization interval allows quasi continuous speech reception [KUN 05]. In addition, short packetization interval (and consequently shorter packets) causes less problems for voice quality reception if a packet is lost [KUN 05]. However, using shorter packetization interval implies more bandwidth consumed as it requires more datagrams and thus more control overhead. This overhead becomes more important and should be carefully considered in wireless networks where the scarce bandwidth is limited and more expensive. On the other hand, using longer packetization interval reduces the overall RTP/UDP/IP/MAC overhead introduced by the different headers. Multiple voice frames can be packed into single packet for transmission reducing the number of packets in the network.

To understand the impact of packetization**,** let us consider a voice stream generated by a G.711 codec. If RTP packets are sent every sample, i.e., a packet every 0.125 ms, the total RTP/UDP/IP overhead is 0.320*8000 = 2560 Kbps, which is 40 times the bandwidth required by the codec. The total required bandwidth in this case is 2624 Kbps. However, if RTP packets are sent every 10 ms, the total overhead is reduced to 0.320*100 = 32 Kbps. If RTP packets are sent every 20 ms, the total overhead is further down to 0.320*50 = 16 Kbps. The total required bandwidth in this case is 80 Kbps.

This example shows clearly the bandwidth conservation when using longer packetization interval. This bandwidth conservation is achieved at the expense of longer reconstruction delay as packets contain more speech bytes and the decoder have to wait for longer time before reconstructing the original speech stream (cf. section 4.3.2 for details). Furthermore, packet loss in this case has drastic consequences on the speech quality. When a packet is dropped, then an observable “blank” in the conversation is realized.

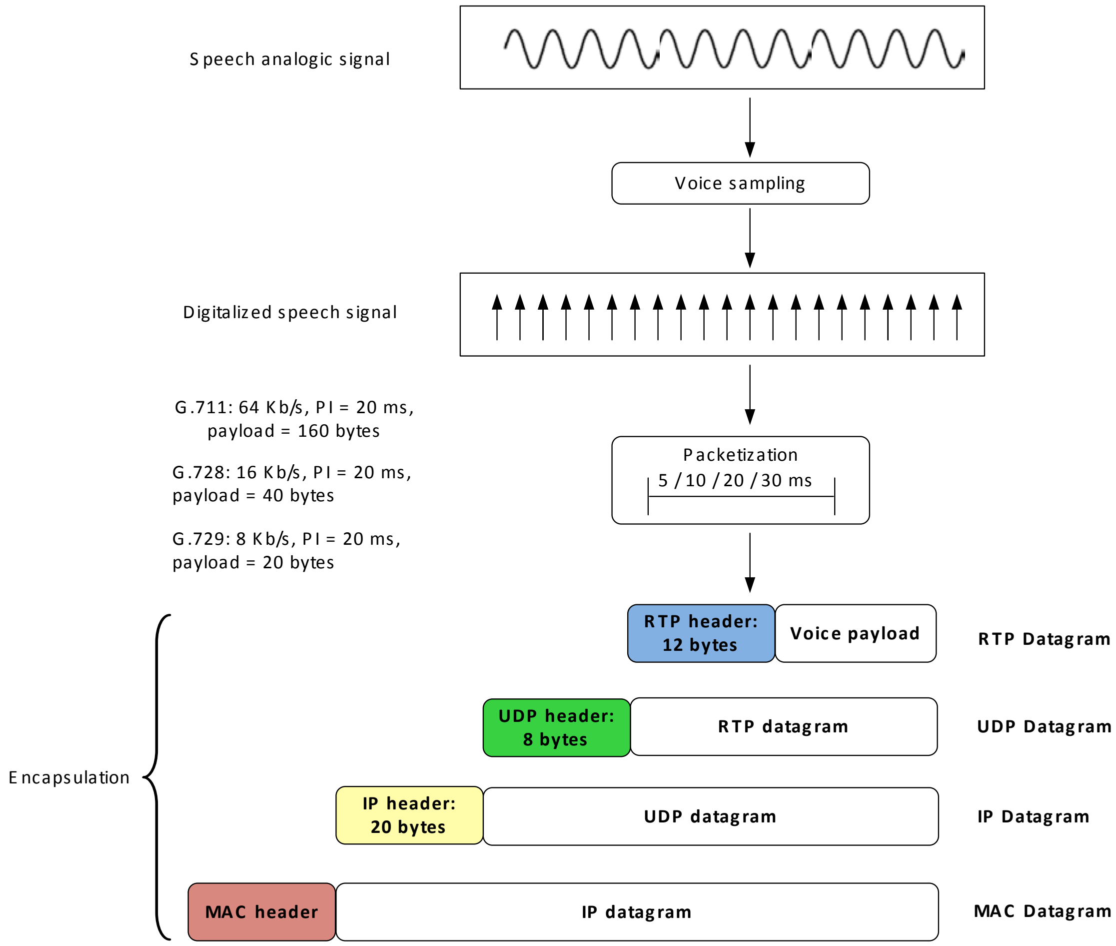

Figure 4-2. Packetization and encapsulation of voice streams.

Voice communication equipments vendors can decide how many speech samples they want to send in one packet [BAH 06]. The packetization intervals recommended are 10, 20, or 30 ms. The default packetization interval of different voice codecs and the corresponding packet payload size as reported in RFC 3551 [SC 03] are shown in Table 4-3.

Table 4-3. Default packetization interval and overhead of different voice codecs.

Codec	Data rate (Kbps)	Default packetization interval (ms)	Payload size (bytes)	IP Packet size (bytes)	Total consumed bandwidth (Kbps)
G.711	64	20	160	200	80
G.722	48	30	150	190	58.56
G.723.1 - 6.3 Kbps	6.3	30	24	64	16.86
G.723.1 - 5.3 Kbps	5.3	30	20	60	15.8
G.726 - 40 Kbps	40	20	100	140	56
G.726 - 32 Kbps	32	20	80	120	48
G.726 - 24 Kbps	24	20	60	100	40
G.726 - 16 Kbps	16	20	40	80	32
G.728	16	20	40	80	32
G.729	8	20	20	60	24
GSM FR	13	20	33	73	29
GSM EFR	12.2	20	31	71	28.2
iLBC 20 ms	15.2	20	38	78	31.2
iLBC 30 ms	13.33	30	50	90	23.89

Notice that the consumed bandwidth in Table 4-3 does not include Layer 2 (data link layer) headers. It includes headers from Layer 3 (network layer) and above only. Therefore, the same codec can consume different amounts of bandwidth depending on which data link layer is used.

4.2.3 Voice decoding sub-system

When the received speech traffic is de-packetized, the reverse process of decoding is performed to reproduce the original signal. Decoding is achieved using the same algorithm as used by the coding sub-system. The coding sub-system and the decoding sub-system may support several different codecs. Therefore, when a voice conversation is initiated, the caller and the receiver negotiate and agree on the voice codec they will use during the conversation. If a waveform coding (such as G.711 codec) is used, the decoder reconstructs the original speech from the amplitude of received Pulse Code Modulated samples. If a Linear Predictive Coding algorithm is used (such as G.729 codec), the decoder uses the parameters received from the sender in the same internal model, which is a replica of the model used at the sender, to synthesize the speaker voice. The speech produced by the decoder in this case, is not the original speech signal reconstructed based on its digital representation, but it is speech locally manufactured based on the received model parameters [KUN 05].

4.2.4 Communication sub-system

The communication sub-system is considered as the core of the global network architecture. Its role is to ensure the reliable transport of voice and data traffics between end-systems. It is composed of the protocol stack where each protocol has a specific task. The different layers of the communication sub-system are shown in Figure 4-3. Each layer has its associated functions and responsibilities, and adds its own control information. The different layers composing the communication sub-system should take into account the characteristics of both the voice traffic, and the limited radio resources of the wireless ad-hoc network.

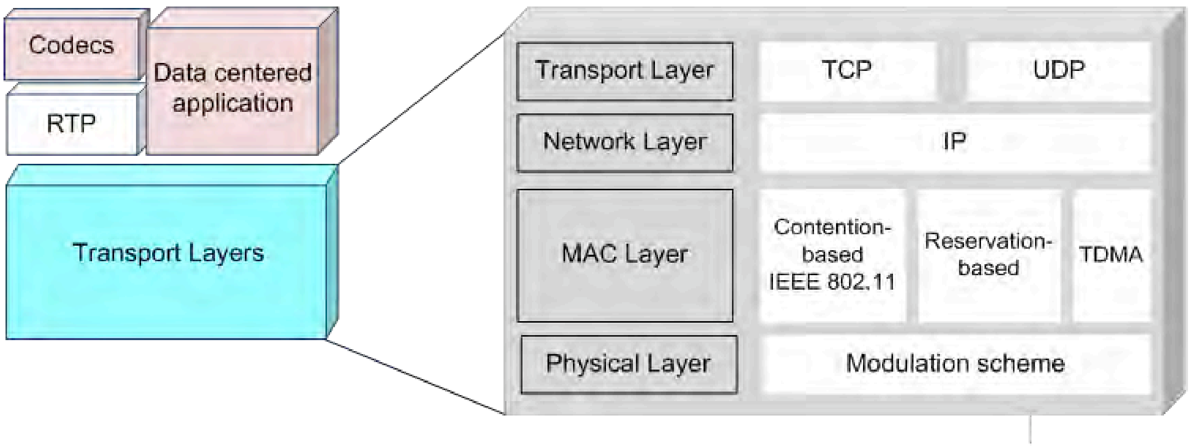

Figure 4-3. Protocol stack for voice/data traffic transmission over wireless ad-hoc networks.

The physical layer is the lowest layer, and represents the direct interface to the wireless medium. The physical layer is responsible of the transmission of the bit stream over the wireless channel. It deals with the electrical specification of the wireless interface hardware, coding, and modulation of digital information into electromagnetic signals. The modulation techniques used in wireless networks are OFDM (*Orthogonal Frequency Division Modulation*), DSSS (*Direct Sequence Spread Spectrum*), and FHSS (*Frequency Hopping Spread Spectrum)*.

The physical layer is out of the scope of our work. The only parameter of the physical layer which is considered in the design of our solution is the channel bit rate. In order to provide a fair comparison with the IEEE 802.11 in performance evaluation, we consider the data rates offered by the different versions of the IEEE 802.11:

- 2 Mbps with the legacy version with FHSS modulation;
- 11 Mbps with the 802.11b version with DSSS modulation;
- 54 Mbps with 802.11g with OFDM modulation.

At the MAC sub-layer, we suppose the use of our reservation protocol presented in this chapter.

The network layer is responsible of routing of data packets from a source to a destination which are more than one hop away from each other. Routing is achieved by a routing protocol which is responsible of end-to-end route discovery. Another function of the network layer is node addressing and packet forwarding. Each packet includes the information necessary for its transmission and forwarding such as the destination address and the next hop node address. The network layer is also responsible of finding an end-to-end path that meets a required bandwidth for a traffic flow, and reserving resources along this path in cooperation with the MAC sub-layer. Routing and end-to-end resource reservation will be the focus of Chapter 7.

The next layer, the transport layer, provides the upper layers with a network interface to the lower layers. Two main protocols have been defined by the IETF [IETF] in this layer: TCP (Transmission Control Protocol) [TCP] and UDP (User Datagram Protocol) [UDP]. TCP is a reliable connection-oriented protocol where packets belonging to a session flow are received at the destination in the same order as they were sent by the source. TCP ensures end-to-end packet delivery, and end-to-end error recovery through acknowledgements and packet retransmission. The other functions provided by TCP are congestion control, and segmentation and reassembly of messages. UDP is an unreliable connectionless protocol. It does not offer flow control, error recovery, and sequencing functions. UDP is used by real-time applications such as voice/video transmission where fast packet delivery is more important than the accuracy of packet delivery. Real-time applications use UDP instead of TCP because TCP is too heavy for these applications. The enhanced functionalities of TCP work well for transmitting large amounts of data, but they are not efficient for real-time media communications [MMS 07].

UDP has no control over the order in which packets arrive at the destination or how long they take to get there. Both of these features are very important to overall voice quality. In order to solve this problem, the RTP protocol is used on top of UDP to ensure ordering functionality, and to control voice packets delay and jitter. The RTP header (cf. Figure 4-4) includes fields that help in providing synchronization and reordering voice packets. The “Payload Type” field is of seven bits and indicates the type of traffic transported by the packet (voice, video…) and the codec type. The “Sequence Number” field is used for reordering arriving packets at the receiver, and the detection of lost packets. It is 16 bits length, and is incremented by one for each transmitted packet. It allows the insertion of out of order packets at the right rank in the reception buffer, and the detection of lost packets. The “RTP Time Stamp” field indicates the instant at which the first byte of the RTP packet was sampled. It allows controlling the delay and delay jitter of voice packets before their delivery to the application [PUG 05]. Table 4-4 gives examples of values taken by the “Payload Type” field.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|V=2|P|X|   CC  |M|     PT      |       sequence number         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           timestamp                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           synchronization source (SSRC) identifier            |
+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
|            contributing source (CSRC) identifiers             |
|                             ....                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4-4. RTP packet header format.

Table 4-4. Examples of the values of “Payload Type” field [PUG 05].

Application	Payload Type value	Codec type
Audio	0	G.711 (μ-law)
	4	G.711 (a-law)
	8	G.723
	9	G.722
	15	G.728
	18	G.729
Video	35	H.261
	34	H.263

4.3 Quality of Service (QoS) for voice traffic

The challenge in voice communication is related to the manner in which human beings communicate. In order to understand each other, the speech signal produced and exchanged between two people must arrive in the same form that it was sent, and within bounded delay. However, when voice packets traverse a network, they may arrive to the destination in an order which is different from the order of transmission. Some packets may take more time to arrive than others, while some packets may not arrive at all. Loss of voice packets causes missing of short segments of speech (phonemes or words), and large delay between the transmission time of a packet and its reception means that the communication is not real-time. Thus, if packets are too much delayed or are lost frequently, there will be significant degradation of the quality of the voice communication. In such conditions, it is difficult for the two communicating entities to continue the conversation.

4.3.1 Quality of Service for voice

The quality of voice communication can be defined from two points of view: QoS experienced by the end user (*Perceived Quality of Service*) and the QoS from the point of view of the network (*Intrinsic Quality of Service*) [HAR 03]. From the end user perspective, QoS is the end user perception of the quality that he/she receives from the network. It determines whether a user is comfortable with the service delivered by the network, and the degree of his/her satisfaction. Besides the qualitative description given by the user, like “quite good” or “very bad”, there is a numerical method of expressing voice quality, which is Mean Opinion Score (MOS) (cf. Table 4-1). The MOS gives a subjective indication of the perceived quality of the speech received after being compressed and transmitted. These are some factors that influence the user perception of the quality of a voice conversation [HAR 03]:

- Clearness and naturalness of listened speech;
- The rhythms, inflection, and cadence of the speech conversation;
- The delay and frequency of “blanks” during the communication;
- The frequency of blocking of the communication;
- The echo due to the speech signal reflected back to the speaker;
- The ability to recognize and differentiate speakers from the received speech.

From the network point of view, the term QoS refers to the network capability to provide the QoS perceived by the end user [KUN 05]. The quality of voice communication from the network perspective is mainly measured by the following QoS parameters: end-to-end transmission delay, delay variation (or jitter), and packet dropping rate. Network protocols try to provide QoS mechanisms that control these parameters without reference to user perception of speech quality. Nonetheless, these mechanisms directly affect the Quality of Service perceived by the end user.

4.3.2 End-to-end transmission delay

Also referred to as end-to-end latency, the end-to-end transmission delay is a critical parameter impacting the quality of voice traffic. It is defined as the finite amount of time experienced by a packet to reach the receiving endpoint after being transmitted from the sending endpoint. In the case of voice, it is equal to the amount of time that it takes for speech to leave the speaker mouth and reach the listener ear. The end-to-end delay is composed of two components: fixed delay component and variable delay component [BAH 06].

The fixed delay component includes the three following delays:

- The coding (or sampling) and decoding delay;
- Packetization/De-packetization delay;
- Serialization delay;

Coding/Decoding delay: This delay is due to the encoder at the voice source and decoder at the voice destination. It is caused by all converting operations: original voice stream from the analog signal to a digital one, received digital signal to a voice signal. This delay depends on the characteristics of the voice codec, and varies from a few milliseconds with G.711 codec to more than 50 ms with G.723.1.

Packetization/De-packetization delay: As illustrated in section 4.2.2, the digitized speech signal is transmitted in the form of packets. The packetization delay is the delay due to putting the samples into packets and placing headers. Thus, speech samples are buffered, and a packet is formed only when the packet payload is filled. The packetization depends on the packetization interval and the number of voice samples that compose one voice packet [BAH 06]. For example with G.711 codec, a packet includes 160 PCM samples of eight bits (i.e., 160 bytes). The delay that suffers the speech bits due to packetization in this case is:

$$\text{G.711 Packetization delay} = \frac{160 \times 8\,bits}{64Kb/s} = 20\,ms \tag{4-5}$$

The de-packetization at the receiving end is nearly the same amount of delay. The total delay for packetization and de-packetization is therefore 40 ms.

Serialization delay: The serialization delay is the time required to place a bit or byte onto a wireless interface. It is directly related to the clocking rate of the wireless interface [CIS 06-2]. This delay is negligible compared to the other delay components.

The variable delay component includes the propagation delay, and the transport or queuing delay:

Propagation delay: The propagation delay is the duration taken by a signal to travel through the wireless channel. The propagation delay depends on the distance between two communicating nodes, and is in the order of a few milliseconds. Although this delay is almost negligible compared to the other components of the end-to-end delay, propagation delay in conjunction with queueing delays can cause noticeable speech degradation.

Queuing delay: The queuing delay is the duration of traversing intermediate wireless nodes between the source and the destination. When packets arrive at an intermediate node, they are temporarily stored in buffers. Packets are queued for three reasons. The first reason is packet processing. The second reason is the lack of resources availability at the MAC sub-layer. When a packet is received from the upper layers, it is queued at the MAC sub-layer until the packet is successfully transmitted. The factor which has an impact on the queuing delay at the MAC sub-layer of a node is the number of contending nodes to access the shared medium. The third reason is the waiting for route discovery at the network layer.

The other factors which have an impact on the total queuing delay are the buffers size and the packet queuing algorithm. Most popular queuing techniques are FIFO (First In First Out), and WFQ (Weighted Fair Queuing), and Round Robin (RR).

The total queuing delay on a path depends on the number of intermediate nodes on the path, and the traffic load in their neighborhood. In a congested network, queuing delay can add up to many seconds of delay.

The G.114 [ITU 96-2] and G.131 ITU-T recommendations [ITU 03] suggest that for an acceptable human interaction, the speech end-to-end delay should be less than 300 ms. 150 ms is the upper bound of end-to-end delay in order to achieve good quality speech reception.

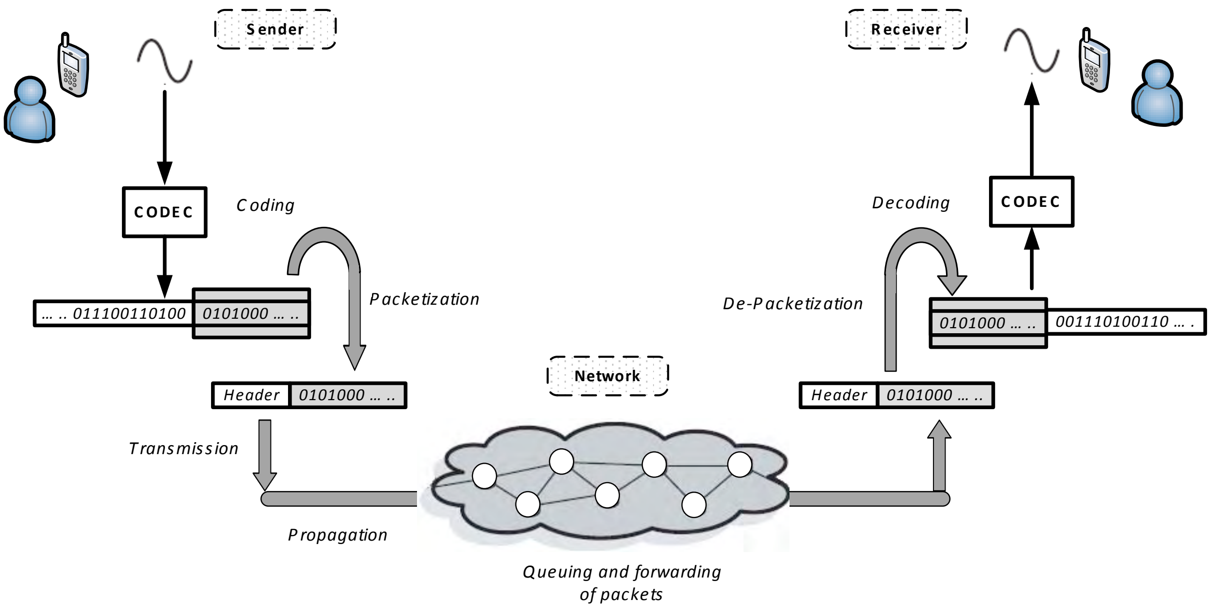

Figure 4-5. Different components of the end-to-end delay.

The queuing delay is the most important part in the end-to-end delay. While once the communication components are fixed, there is no solution to reduce the fixed delay component, QoS solutions focus on reducing the queuing delay at the different layers. However, despite the lot of research devoted to propose techniques that reduce the queuing delay, this remains a big challenge.

4.3.3 End-to-end delay variation or jitter

Jitter is defined as the variation of packet interarrival time [CIS 06]. While the sender is expected to send voice packets at a regular interval (for example, one packet every 20 ms), these packets may not arrive at the receiver at that same regular interval. This jitter is the result of the fact that voice packets do not cross the network at the same speed. One source of jitter is variation of traffic load in the network during different periods of time. Hence, an increase of jitter may be a consequence of transient congestion on the network. The latter cannot transport voice packets in a constant time. Another source of jitter is the network dynamics. In highly dynamic network, voice packets travel by different paths with different lengths causing a variation in their end-to-end transmission delay.

The effects of jitter on speech reception quality can be reduced through storing voice packets in a jitter play-out buffer at the receiver before delivering them to the decoder. While smooth jitter can be obtained through this technique, it causes an undesirable increase of delay.

The value of the jitter ranges from a few milliseconds to tens of milliseconds. For a good quality speech reception, jitter must be less than 20 ms [ACC].

4.3.4 Packet dropping rate

Too many lost packets prevent the receiver from completely reconstructing the original sent voice stream, and gaps in the received speech are detected. These gaps correspond to words or sentences which are completely missed.

Unlike data traffic where retransmission of packets is used to overcome these packet loss problems, the real-time nature of voice traffic makes it impractical the retransmission of lost voice packets. Even if retransmitted, a voice packet arrives too late, and thus is no more useful in the process of the original voice stream reconstruction. Moreover, retransmission of a voice packet after transmission failure may contribute in increasing the queuing delay of new generated voice packets which have more chance to arrive on time. This fact poses more challenge on the communication sub-system which should avoid as far as possible discarding voice packets.

The packet dropping rate of voice packets in MANETs depends on the channel quality and the frequency of network topology changes. Table 4-5 summarizes the consequences of the values of the QoS parameters on the service level of speech reception [HAR 03].

Table 4-5. Recommended QoS parameters values for different voice quality levels.

Service level / Parameter	Good	Average	Bad
End-to-end delay (D)	< 150 ms	150 ms < D < 400 ms	> 400 ms
Jitter (J)	< 20 ms	20 ms < J < 50 ms	> 50 ms
Packet loss rate (R)	< 1%	1% < R < 3%	> 3%

4.4 Voice communication over Wireless Ad-hoc Networks

A MANET infrastructure is able to accommodate voice communication service, if it provides a good level of speech reception in different traffic load conditions, i.e., if it delivers the sound pronounced by the speaker to the listener ear within 150 ms, with 1% maximum packet loss, and with 20 ms maximum jitter, even at high traffic load. However, this level of QoS is very difficult to achieve in MANETs. Actually, the performances of MANETs infrastructures in regard of voice traffic are very poor. These poor performances are mainly due to the inability of MAC protocols to handle the unfavorable characteristics of MANETs (cf. Chapter 1).

As noticed in Chapter 2, MAC protocols control access to the wireless medium, and define how mobile nodes can share the limited wireless bandwidth resource in an efficient manner. They are also responsible of resolving conflicts among different nodes for channel access. Thus, they have a strong impact on the QoS provided to voice traffic in terms of transmission delay, delay jitter, and dropping rate, because they have a direct bearing on how reliably and efficiently voice packets can be transmitted.

A detailed discussion on how MAC protocols affect the QoS provisioning for voice traffic is given below.

- ***Impact of the MAC layer on the transmission delay:*** As noticed in section 4.3.2, the most important component of the end-to-end transmission delay is the queueing delay. A big proportion of the queuing delay is due to the queueing delay at the MAC layer which is called the access delay. When a packet arrives to the MAC layer, the packet is not transmitted on the physical medium immediately. In order to resolve contention between contending nodes, the MAC layer is required to wait a certain amount of time before sending its queued packet waiting for transmission. Hence, the delay incurred by channel access may cause some of the packets waiting for transmission to miss their target delay, which is not acceptable for applications that have stringent delay requirements such as voice.

- ***Impact of the MAC layer on the delay jitter:*** The MAC layer can cause an increase of delay jitter due to the random channel access delay at the MAC layer. The access delay is significantly affected by the traffic load in the network. Low access delays are expected at low traffic load, while an increase of the access delay is observed with the increase of traffic load. Thus, a low jitter can be obtained only if a constant access delay is ensured for all packets of the same traffic flow. Mechanisms that ensure a guaranteed and deterministic periodic access to voice packets are required at the MAC layer in order to provide low jitter.

- ***Impact of the MAC layer on the Dropping rate:*** There are two sources of packet loss in MANETs: (1) network packet losses, mainly due to network congestion (nodes buffer overflow), link failures and re-routing, and (2) discarded packet loss for packets experienced excessive delay or excessive collision at the MAC layer. In the IEEE 802.11, a packet is retransmitted up to a maximum retransmission limit, and it is dropped when the retransmission limit is reached. In some other access schemes (such as [SK 96], [SK 99], [KSS 02], and [KAN 02]) a deadline is associated with each packet, and the packet is dropped by the MAC layer if this deadline is exceeded. Thus, the dropping rate hinges on the probability of transmission failure and the packet collision which rely on the efficiency of the MAC layer in reducing the effect of traffic load.

Regarding these important effects, medium access schemes should provide mechanisms that reduce the effect of packet collision and achieving low access delay in order to meet the stringent QoS requirements of voice traffic.

4.5 Conclusion

In this chapter, we presented the components of voice/data communication over wireless ad-hoc networks architecture. First, we exposed the characteristics of each one of these components. Then, we presented the basic concepts and characteristics of voice traffic with more emphasis on their QoS requirements. After, we pointed out the important role that plays the MAC sub-layer in QoS provisioning for voice traffic support in MANETs.

Existing reservation-based MAC protocols include some features which are interesting for voice traffic such as the low contention rate, and low access delay, and low impact of traffic load. However, despite these interesting features, there are still several challenges that make them not able to provide QoS for voice. For example, contention is still present, at least during the reservation request phase. Voice sources may contend and wait for long time before getting periodic contention-free access. Thus, the impact of contention during the reservation request phase on the voice traffic performance should be carefully considered in the design of a reservation protocol.

Another issue rises with voice activity detection (cf. section 4.2.1.4). When a voice traffic source does not generate traffic, its reserved bandwidth remains unused leading to a significant waste of bandwidth.

In the following chapter, we propose extensions to the reservation protocol proposed in chapter 3 in order to support voice traffic. Thus, we adapt it so that it takes into account the characteristics of voice traffic, and satisfy its QoS requirements.

Chapter 5: Bandwidth Allocation Scheme for Voice Traffic Support over MANETs

5.1 Introduction

The reservation protocol presented in the chapter 3 represents a basic reservation scheme that can be used to provide QoS for a variety of multimedia and real-time applications. Among these applications, we are interested in QoS provisioning for voice traffic, while providing an acceptable level of QoS for data traffic. A typical voice/data communication application that may use our protocol involves an Ad-hoc Network composed of a number of mobile nodes (which can be laptops, PDAs, or mobile phones) equipped with wireless interfaces. Applications that run over these mobile devices are either voice oriented applications such as Skype or Messenger for voice communication (which are out of the scope of our work), or data centric applications such as e-mail exchange and web navigation. Thus, any pair of nodes can use our protocol for voice or data communication.

In order to be used for voice/data traffics integration and ensure the required QoS of voice traffic, several protocol parameters need to be analyzed. This chapter addresses this parameterization issue which consists in answering the following questions:

- What are the parameters and enhancement schemes that give the best performances for both voice and data traffics?

- What is the impact of the slot length on bandwidth utilization, and what is the best choice of the slot length?

- What are the solutions for contention resolution during the *RSF*?

- How can voice and data traffic sources share efficiently the limited bandwidth?

We will present basic solutions to resolve these issues.

5.2 Adaptive Reservation Protocol for Voice Traffic Support over MANETs (ARPV)

Our protocol called Adaptive Reservation Protocol for Voice Traffic Support over MANETs (ARPV) includes methods for bandwidth reservation for voice/data traffics integration in Wireless Ad-Hoc Networks. Therefore, the aim of the protocol is to control the sharing of bandwidth between voice and data sources. This objective brings to the several questions that should be considered:

- What is the best choice of the slot length given the characteristics of voice codecs?

- How can voice sources reserve and share bandwidth with data sources knowing that they are not active all the time.

- How to resolve contention during the Reservation Sub-Frame (*RSF*).

5.2.1 Issue 1: Optimal slot length selection

The purpose of this part is to investigate different choices of the slot length, their advantages and drawbacks, and their impact on the efficiency of bandwidth utilization. As noticed in section 3.3.3, the data slot length *(L_{slot})* is set to the transmission time of one real-time packet with the different layers (RTP, TCP/UDP, IP, and MAC headers) overheads. In this chapter, it corresponds to the packets generated by a voice traffic source.

We consider that voice sources may use different codecs. Therefore, as voice codecs have not the same packet length (cf. section 4.2.2), the selection of the data slot length is not an obvious task, and should be carefully studied. A slot length selection may be suitable for some codec, but not practicable for other codecs.

Our primary focus in this section is to determine the impact of the slot length on the performance of our reservation protocol. The efficiency of a solution for slot length selection is assessed according to the following metrics: fragmentation overhead, and the number of data slots per super-frame. Since the fragmentation incurs some control overhead, it is preferable to send a single voice packet as a whole without fragmentation. The number of data slots per super-frame is important because it has a direct impact on the number of voice connections that can be supported by the super-frame.

In the rest of this section, we provide a theoretical analysis of the different choices. Experimental analysis is given in other sections. Thus, in order to simplify the theoretical analysis, we consider that the super-frame is composed only of data slots. The reservation sub-frame is considered null.

Before detailing the different potential solutions, let us present the basic structure of data, control, and acknowledgement frames.

5.2.1.1 Solution 1 for “optimal slot length selection” issue

The first choice of the slot length consists in choosing the packet length of the codec that have the longest packet length among the considered codecs (cf. section 4.2.1) as the data slot payload, i.e., 160 bytes with G.711. In this case, the data slot length (in bytes) including the different headers (from the RTP up to the PHY layer) and MAC layer acknowledgement, is 256 bytes.

We calculated the slot length and the number of data slots per super-frame in this case for different channel bit rates as shown in Table 5-1. We consider the three channel bit rates defined in the different versions of the IEEE 802.11 standard, i.e., 2 Mb/s, 11 Mb/s, and 54 Mb/s. Notice that the number of data slots per super-frame is approximate, and is given without considering the bandwidth consumed by the reservation sub-frame.

Table 5-1. The data slot length and the number of data slots per super-frame when the data-slot corresponds to the payload of G.711.

Channel bit rate	Data slot and ACK mini-slot length	Number of slots per super-frame (approximately)	Number of voice connections supported by the super-frame (approximately)
2 Mb/s	1.024 ms	20	20
11 Mb/s	0.184 ms	108	108
54 Mb/s	0.037 ms	530	530

The advantage of this solution is that it avoids the overhead due to fragmentation. For instance, the data slot is long enough to accommodate packets of all considered codecs. Thus, each voice connection reserves one slot, and there is no need of fragmentation. The major drawback of this solution is that bandwidth is not used efficiently with voice codecs which are characterized by short packets. For example, if a G.729 connection reserves a slot, more than 50% of the slot is wasted as it is not used for transmission.

5.2.1.2 Solution 2 for "optimal slot length selection" issue

The second way to choose the slot length consists in choosing the data slot payload that corresponds to the packet length of the codec with the shortest packet length among the considered codecs, i.e., 20 bytes with G.729 (cf. Table 4-3). In this case, voice sources that use a codec with packet length longer than the data slot payload are required to fragment their voice packets, and reserve several slots per super-frame. Fragmented packets should have the "More Fragment" bit set at "1". These packets are re-assembled at reception before delivering them to the upper layer.

We calculated the number of voice connections and the number of slots required for different codecs (shown in Table 5-3). As shown in the table, the number of voice connections supported by the super-frame with this solution depends on the codec used. For instance, it is equal to the number of slots per super-frame if voice connections use G.729 codec. It may be much lower if the used codec have packet length longer than the data slot length.

Table 5-2. The data slot length and the number of slots per super-frame when the data-slot corresponds to the payload of G.729.

Channel bit rate	Data slot and ACK mini-slot length	Number of slots per super-frame (approximately)
2 Mb/s	0.464 ms	43
11 Mb/s	0.082 ms	243
54 Mb/s	0.016 ms	1184

Table 5-3. The number of voice connections supported by the super-frame when the data-slot corresponds to the payload of G.729.

Channel bit rate	Codec	Number of slots required	Number of voice connections supported (approximately)
2 Mb/s	G.711	8	5
	G.726 - 40 Kbps	5	8
	G.726 - 32 Kbps	4	10
	G.726 - 24 Kbps	3	14
	G.726 - 16 Kbps	2	21
	G.728	2	21
	GSM FR	2	21
	iLBC 20 ms	2	21
11 Mb/s	G.711	8	30
	G.726 - 40 Kbps	5	48
	G.726 - 32 Kbps	4	60
	G.726 - 24 Kbps	3	81
	G.726 - 16 Kbps	2	120
	G.728	2	120
	GSM FR	2	120
	iLBC 20 ms	2	120
54 Mb/s	G.711	8	148
	G.726 - 40 Kbps	5	236
	G.726 - 32 Kbps	4	296
	G.726 - 24 Kbps	3	393
	G.726 - 16 Kbps	2	590
	G.728	2	590
	GSM FR	2	590
	iLBC 20 ms	2	590

The advantage of this solution is that it avoids the waste of bandwidth due to partially used slots, i.e., slots which are not completely used for transmission. Its major drawback is the high fragmentation overhead for codecs with long packets. From Table 5-1 and Table 5-3, we can see that the 1st solution is better than the 2nd one in terms of the number of supported connections when G.711 and G.726 codecs are used. This is mainly because the 1st solution is more suitable for these two codecs which have relatively long packets compared to the other codecs, i.e., G.728, GSM FR, and iLBC. In addition, with the 2nd solution these two codecs consume more bandwidth due to the fragmentation overhead. However, with the 2nd solution, more voice connections are supported than the 1st solution for G.728, GSM FR, and iLBC codecs.

5.2.1.3 Solution 3 for "optimal slot length selection" issue

The third potential solution that we consider is an enhancement of the 2nd solution. In order to avoid the waste of bandwidth due to fragmentation overheads, we propose that a voice connection that needs several slots do not fragment its packets. Thus, the voice connection reserves several contiguous slots, and sends its voice packets without fragmentation. Consequently, the packet header is sent only once at the beginning of the first reserved slot. In order to see the theoretical bandwidth gain of this solution compared to solution 2, we compute the number of slots required for each codec and the number of connections supported by the super-frame (cf. Table 5-4).

Table 5-4. The number of slots required by voice codecs and the number of voice connections supported by the super-frame with the 3rd solution.

Channel bit rate	Codec	Number of slots required	Number of voice connections supported (approximately)
2 Mb/s	G.711	3	14
	G.726 - 40 Kbps	2	21
	G.726 - 32 Kbps	2	21
	G.726 - 24 Kbps	2	21
	G.726 - 16 Kbps	2	21
	G.728	2	21
	GSM FR	2	21
	iLBC 20 ms	2	21
11 Mb/s	G.711	3	81
	G.726 - 40 Kbps	2	121
	G.726 - 32 Kbps	2	121
	G.726 - 24 Kbps	2	121
	G.726 - 16 Kbps	2	121
	G.728	2	121
	GSM FR	2	121
	iLBC 20 ms	2	121
54 Mb/s	G.711	3	393
	G.726 - 40 Kbps	2	590
	G.726 - 32 Kbps	2	590
	G.726 - 24 Kbps	2	590
	G.726 - 16 Kbps	2	590
	G.728	2	590
	GSM FR	2	590
	iLBC 20 ms	2	590

Following Table 5-4, codecs with long packets consume less bandwidth compared to the 2nd solution, thanks to the low fragmentation overhead. For instance, G.729 connections consume only one slot per super-frame, G.711 connections consume three slots (instead of 8 with the 2nd solution), and all the other codecs consume two slots. As a consequence, this solution gives better performance in terms of number of supported voice connections than the 2nd solution. However, the efficiency of this solution is constrained by the availability of a set of contiguous slots. For instance, a G.711 source is able to send its voice packets without fragmentation only if there are three contiguous slots available for reservation.

5.2.1.4 Conclusion

Table 5-5 summarizes the different solutions considered for the optimal slot length selection issue.

Table 5-5. Potential solutions for the slot length choice.

Solutions	Description	Advantages	Drawbacks
Solution 1	The slot length corresponds to the packet size of G.711 codec.	No need of fragmentation. Data slot long enough so that it carries packets of all considered codecs.	Not suitable for codecs with small packets. Data slots are not fully used.
Solution 2	The slot length corresponds to the packet size of G.729 codec.	Data slots are fully used.	High fragmentation overhead for codecs with long packets.
Solution 3	Enhancement of solution 2. Sources with long packets reserve contiguous slots if available.	Prevents fragmentation overhead of solution 2.	Fragmentation overhead avoided only if contiguous slots are available.

5.2.2 Issue 2: Time-slot reservation for voice

The purpose of this part is to propose solutions to the following issue: given that voice traffic sources are not active for all the duration of the connection due to the Voice Activity Detection (cf. section 4.2.1.4), how can voice and data sources share the limited bandwidth resource without causing a waste of bandwidth?

As illustrated in section 4.2.1.4, voice sources are not active during all the connection. They are equipped with a Voice Activity Detector (VAD), and follow an alternating pattern of talkspurts and silence periods (On/Off). A simple solution for bandwidth reservation for voice traffic sources with this behavior consists in making voice sources reserve slots for all the duration of the connection, and releasing them at the end of the connection. Data traffic sources are allowed only to use slots which are not reserved by voice traffic sources. While simple, this solution is inefficient in terms of bandwidth utilization because slots reserved by voice sources are not used for transmission during idle periods. In the rest of this section, we propose a solution to reduce this waste of bandwidth.

5.2.2.1 Our solution

In this section, we extend the protocol proposed in chapter 3 in order to take into account the characteristics of voice traffic.

One way to avoid the waste of bandwidth due to silence periods consists in recycling slots which are not used for transmission during silence periods so that other traffic sources can use them. Indeed, the main idea of our solution can be summarized in the following points:

- Reserve a slot to each voice traffic source for the duration of the connection.
- A slot reserved for a voice traffic source can be temporarily released when the voice traffic source goes to the sleep mode.

- Neighboring data sources can use temporarily released slots for packet transmission.
- Slots are temporarily released at both sender and receiver neighbors so that the sender neighbors can use these slots for data packets reception, and receiver neighbors can use them for data packets transmission.
- The voice source has the opportunity to restore its released slot when it wakes-up.

In order to achieve this multiplexing between voice and data sources, we need to consider the following slot status variables:

- *Available for Transmission slots set (AT_i)*: The set of slots available for transmission at node *i*. A slot is considered available for transmission if no neighbor has reserved the slot for reception.
- *Available for Reception slots set (AR_i)*: The set of slots available for reception. A slot is considered available for reception if no neighbor has reserved the slot for transmission.
- *Reserved for Transmission slots set (RT_i):* The set of slots reserved for transmission. One or more neighbors have reserved the slot for packet transmission. The slot cannot be reserved for reception.
- *Reserved for Reception slots set (RR_i):* one or more neighbors have reserved the slot for packet reception, and the slot cannot be reserved for transmission.
- *Temporarily Transmission Released slots set (TTR_i):* the slot is reserved by a voice source, but is temporarily released because the voice source is in the silence phase. This slot can be used for reception by neighboring data sources.
- *Temporarily Reception Released slots set (TRR_i):* the slot is reserved by a voice connection receiver, but is temporarily released because the voice source is in the idle phase. Neighboring data sources can use this slot for transmission until the voice connection wakes-up.

In addition to these slot state variables, each node *i* needs to record the set of its reserved slots for transmission and reception, and the set of its temporarily released slots. Thus, we need the following variables:

- *My Transmission Reserved slots* (MTR_i): the set of slots reserved by the node for transmission.
- *My Reception Reserved slots* (MRR_i): the set of slots reserved by the node for reception.
- *My Temporarily Transmission Released ($MTTR_i$):* the set of slots temporarily released by the node for transmission.
- *My Temporarily Reception Released ($MTRR_i$):* the set of slots temporarily released by the node for reception.

Initially, all slots are available for transmission and reception. The RT_i set and the RR_i set are updated each time a slot is reserved by neighbor nodes for transmission and reception respectively. MTR_i and MRR_i are updated each time a slot is reserved or released by the node for transmission or reception respectively.

When a new voice connection is initiated, the voice source node tries to reserve the required number of slots (depending on the used codec) with the intended receiver. Slots that can be reserved by the sender are slots which are not reserved for reception, and slots that can be reserved for reception are slots which are not in the *RT* status or *TTR* status. Thus, since a slot is dedicated to a voice connection during all its life time, slots temporarily released can be used only by neighboring data sources, and cannot be reserved for a new voice connection.

Once a voice connection has reserved successfully the required slots, it starts sending its voice packets on its reserved slots. Thus, during the activity period, the voice source generates and sends one voice packet per super-frame. When the voice source switches to the silence mode, it releases temporarily its reserved slots by leaving them empty. When neighbor nodes detect clear channel during these slots, they mark them as temporarily transmission-released and are allowed to use them for data reception. Similarly, a slot is considered temporarily reception released by the receiver neighbors, when idle channel is detected during the ACK mini-slot of the corresponding slot.

A voice source in the *Off* state checks permanently its packet queue. When packets are present in the queue, the MAC sub-layer of the source node concludes that the voice source switches to the activity period, and temporarily released slots must be restored. The MAC sub-layer executes the collision resolution scheme during the *RSF*, and follows the five reservation phases in order to restore its reservation with the intended receiver. In the mean time, voice packets waiting for reservation are queued until the temporarily released slot is restored. In order to limit the queueing delay of voice packets, we consider that a voice packet waiting for transmission should be transmitted before the arrival of the next one. Thus, if a new voice packet arrives while the previous one is still not transmitted, the old packet is dropped and the new one is kept for transmission.

The reservation steps for a voice traffic source and the different statuses are illustrated in Figure 5-1. The *"Contention"* state denotes the state where the voice source generates a new talkspurt while reservation is not yet established. In this state, the voice source contends to reserve a slot or restore its reserved slot if a slot has been temporarily released. The *"RTS sent"* state designates the situation where a voice source in the *"Contention"* state sends an RTS packet and is waiting for a *CTS* from the intended receiver or CR from one of its neighbors. The voice source returns to the *"Contention"* state if a CR or collision is sensed in the 2nd control mini-slot of the CRS. In the *"RTS sent"* state, if a *CTS* is received, the voice source sends a *ResvRTS* and enters the *"ResvRTS sent"* state. It sends a *ResvConfirm* and enters the *"Reservation"* state if a *ResvCTS* is received, or returns back to the *"Contention"* state otherwise.

In the *"Reservation"* state, the voice source sends its voice packets every super-frame on its reserved slots, and enters the *"Silence"* state if no voice packet is received during a super-frame. In the *"Silence"* state, slots reserved by the voice source are considered temporarily released and can be reserved by neighboring data sources. When a new talkspurt is generated the voice source returns back to the *"Contention"* state in order to restore its reserved slots.

When the voice source is in the *"Contention*", *"RTS sent"* or *"ResvRTS sent"* state and have a queued packet, it drops the queued packet if a new voice packet is generated.

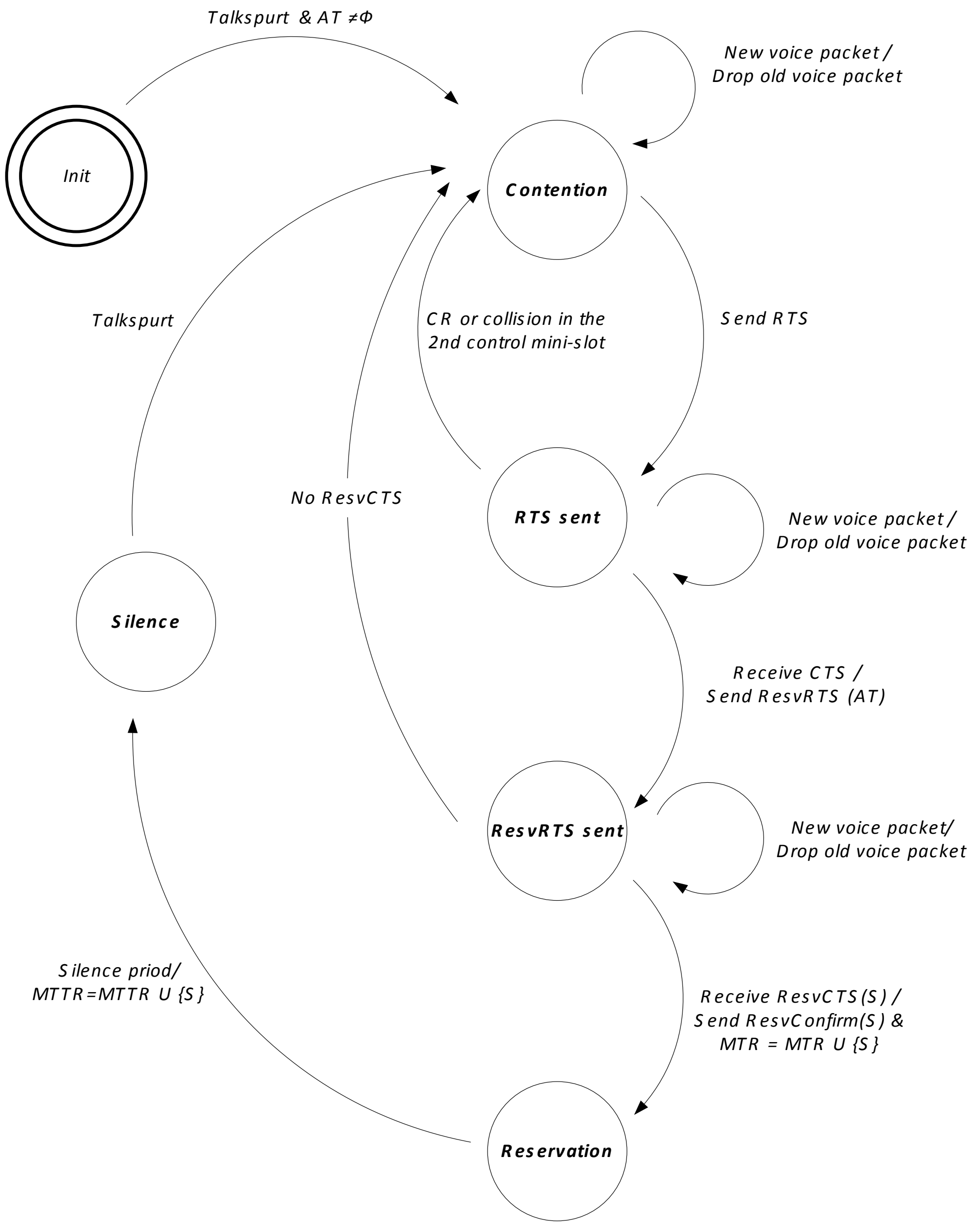

Figure 5-1. State diagram of slot reservation for voice traffic sources.

Figure 5-2 shows the state diagram of a voice traffic receiver. First, the receiver is in the *"Waiting"* state waiting for a reservation request from one of its neighbors. If a collision is sensed in the 1^{st} control mini-slot, the receiver sends a CR packet in the 2^{nd} control mini-slot. If an RTS is received, the receiver sends *CTS* in the 2^{nd} control mini-slot, and enters the *"CTS sent"* state. In the *"CTS sent"* state, the receiver waits for a *ResvRTS* from the voice source. It sends a *ResvCTS* and switches to the *"Reservation"* state if the *ResvRTS* is received. In the *"Reservation"* state, the receiver acknowledges each received voice packet, and returns back to the *"waiting"* state if no voice packet is received during the reserved slot. In the *"waiting"* state, slots reserved for reception are considered temporarily reception released and can be temporarily reserved for transmission by neighboring data sources.

Figure 5-3 shows the transitions and actions executed by idle nodes. Note that these actions are also executed by nodes in the *"Contention"* and *"Reservation"* state in the diagrams of Figure 5-1 and Figure 5-2.

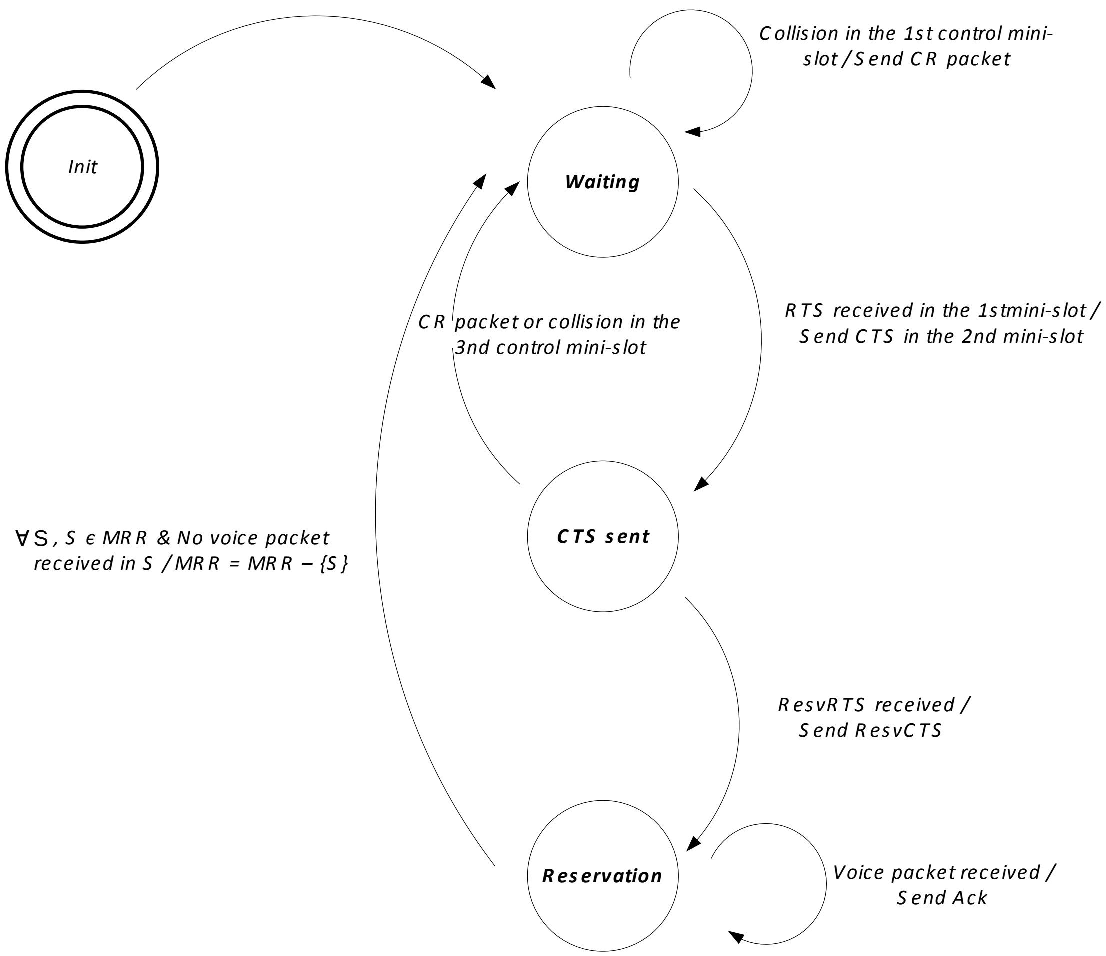

Figure 5-2. State diagram of slot reservation for voice reception.

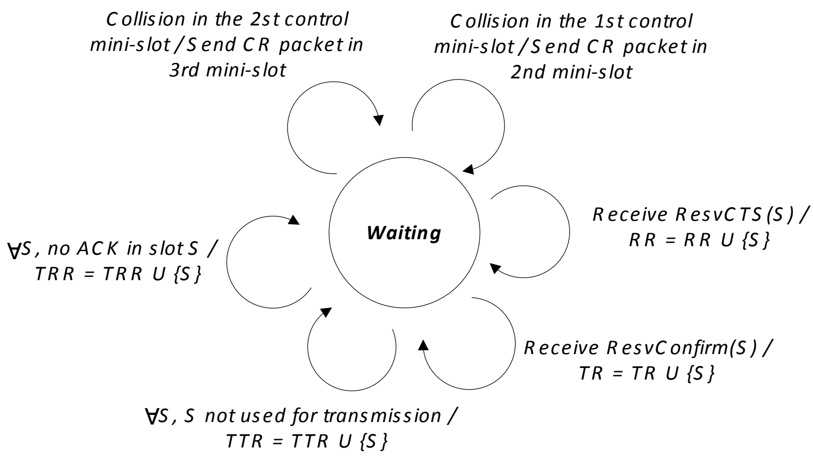

Figure 5-3. State diagram of idle nodes.

5.2.2.1.1 Solution properties

The set of data sources, voice sources, and idle nodes are denoted N^v, N^d, *and* N^{idle} respectively.

The connectivity between nodes forming the network is represented by an $n \times n$ symmetric one-hop connectivity matrix $Neigh_1$ given by:

$$(Neigh_1)_{ij} = \begin{cases} 1, & \text{if nodes } i \text{ and } j \text{ can communicate with each other and } i \neq j \\ 0, & \text{otherwise} \end{cases}$$

At the initialization of the system, the slot status variables are set as follows:

- $AT_i = AR_i = \{1, 2, ..., S\}$
- $RT_i = RR_i = TTR_i = TRR_i = MTR_i = MRR_i = MTTR = MTRR = \phi$

At any moment, the slot status variables should have the following properties:

Property 5-1.

$$\forall i, AR_i \cap RT_i = AR_i \cap MTR_i = AR_i \cap MRR_i = AR_i \cap MTTR_i = AR_i \cap MTRR_i = \phi$$

For a node *i*, a slot reserved for transmission or reception or temporarily released by node *i* cannot be considered available for reception.

Property 5-2.

$$\forall i, AT_i \cap RT_i = AT_i \cap RR_i = AT_i \cap MTR_i = AT_i \cap MRR_i = AT_i \cap MTTR_i = AT_i \cap MTRR_i = \phi$$

For a node *i*, a slot reserved for transmission or reception or temporarily released by node *i* cannot be considered available for transmission.

Property 5-3.

$$\forall i \in N^v, \forall s, s \in MTTR_i \Rightarrow s \in MTR_i$$

For a voice traffic source *i*, a slot temporarily transmission released is considered as reserved for transmission by the voice source *i*.

Property 5-4.

$$\forall i \in N^v, \forall s, s \in MTRR_i \Rightarrow s \in MRR_i$$

For a voice traffic source *i*, a slot temporarily reception released is considered as reserved for reception by the voice source *i*.

Property 5-5.

$$\forall i \in N^v, \forall s, (s \in MRR_i \Rightarrow \exists j \in N^v, (Neigh1)_{ij} = 1 \wedge s \in MTR_j \wedge$$
$$(\forall k \neq j, (Neigh1)_{ik} = 1 \wedge k \in N^v \Rightarrow s \notin MTR_k))$$

For a voice source i, a slot *s* reserved for reception should be marked reserved for transmission by only one neighbor node which is the source.

Property 5-6.

$$\forall i \in N^v, \forall s, (s \in MTRR_i \Rightarrow \exists j \in N^v, (Neigh1)_{ij} = 1 \wedge s \in MTTR_j \wedge$$
$$(\forall k \neq j, (Neigh1)_{ik} = 1 \wedge k \in N^v \Rightarrow s \notin MTR_k))$$

For a voice source i, a slot *s* temporarily reception released should be marked temporarily transmission released by only one neighbor node which is the source.

Property 5-7.

$$\forall i \in N^v, \forall s, (s \in MTR_i \Rightarrow \exists j \in N^v, (Neigh1)_{ij} = 1 \wedge s \in MRR_j \wedge$$
$$(\forall k \neq j, (Neigh1)_{ik} = 1 \wedge k \in N^v \Rightarrow s \notin MRR_k))$$

For a voice source *i*, a slot *s* reserved for transmission should be marked reserved for reception by only one neighbor node which is the destination.

Property 5-8.

$$\forall i \in N^v, \forall s, (s \in MTTR_i \Rightarrow \exists j \in N^v, (Neigh1)_{ij} = 1 \wedge s \in MTRR_j \wedge$$
$$(\forall k \neq j, (Neigh1)_{ik} = 1 \wedge k \in N^v \Rightarrow s \notin MTRR_k))$$

For a voice source *i*, a slot *s* temporarily transmission released should be marked temporarily reception released by only one neighbor node which is the destination.

5.2.2.1.2 End of connection detection

A particular issue with this solution appears when a voice traffic source finishes its transmission. Thus, when the voice source ends its transmission, it is blocked in the *Silence* state, and its reserved slot will remain indefinitely in the temporarily released status. Thus, the slot cannot be reserved by other new voice sources. A solution to this issue consists in associating a *"connection timeout"* with each temporarily released slot. If a voice connection does not restore its temporarily released slot before the *"connection timeout"* expires, the voice source is assumed finished its transmission, and the slot is considered definitively released. The *"connection timeout"* should be long enough to avoid false connections termination detection. Indeed, the *"connection timeout"* should be longer than the duration of idle period of a voice source.

Algorithm 5-1. Slot status update and end of connection detection algorithm.

Parameters:

- *Connection_timeout[i]: connection timeout associated with slot i*
- *Busy(i): returns true if slot i is busy (packet or collision detection), and false otherwise.*

Initialization:

```
For i 1 to S do
    Set connection_timeout[i];

For each super-frame do
    For each slot i do
        If (i∈TTR) then
            If (connection_timeout[i] = 0) then
                AR  = AR ∪ i ;
                TTR = TTR – {i};
                RT = RT – {i};
            Endif
            If (busy(i)) then
                Set connection_timeout[i];
            Endif
        Endif
    Endfor
Endfor
End
```

5.2.3 Issue 3: Channel access for data traffic sources

The purpose of this part is to propose solutions to the following issue: given the reservation scheme for voice traffic described in section 5.2.2, how can data sources use available slots and slots temporarily released by voice sources for transmission.

Depending on whether data sources are allowed to establish reservations or not, we propose two access schemes for data sources: *Reserve Temporarily Released slots (RTR)* scheme and *Contend for Each Packet (CEP)* scheme. With both schemes, we assume that data sources do not have stringent delay requirements, and packets generated by data sources are queued until successfully transmitted. Data packets arriving when the packet queue is full are dropped.

5.2.3.1 Reserve Temporarily-Released slots scheme (RTR)

In this scheme, data sources are allowed to reserve time-slots along with voice sources, with priority given to voice traffic in slots reservation. Data sources are allowed to negotiate reservations only on slots which are not in the *Reserved-for-reception* status, and slots in the *Temporarily-reception-released* status. Since voice traffic has higher priority than data traffic, new voice sources are allowed to grab reservations made by neighboring data sources. The *ResvRTS* sent by a new voice source includes slots which are not reserved for reception by other voice sources, and slots *reserved for reception* by data sources. The receiver can accept a reservation request on slots not reserved for transmission by neighboring voice sources, and slots *reserved for transmission* by neighboring data sources.

A data source that reserved a temporarily released slot loses its reserved slot when the voice source wakes-up and restores its released slot. Thus, when a data source receives a *ResvCTS* which indicates a slot that it has reserved, the data source releases its reserved slot by leaving it empty. Similarly, if the receiver of a data source receives a *ResvConfirm* for a slot which has been locally reserved, it releases its reception slot by stopping sending ACK packet during the ACK mini-slot of the corresponding slot. After, the data source is required to reserve another slot with its intended receiver if available. While waiting for slot reservation, the data source queues its packets until some neighboring voice source switches to the silence mode or finishes its transmission.

Unlike voice sources where slots are temporarily released during inactivity periods, a slot reserved by a data source is considered definitively released (available) if not used for transmission for one super-frame. Thus, data sources are required to reserve slots each time they have new burst of packets, and each time they lose their reservations.

The access scheme for data traffic sources can be summarized in the state diagram of Figure 5-4. Initially, the data source is in the *"Contention"* state after receiving a burst of packets. In the *"Contention"* state, the data source tries to reserve slots among the slots available for transmission (recorded in the *AT* state variable) and slots temporarily reception released (*TRR* variable). When a *ResvCTS* is received, the data source enters the *"Reservation"* state, and the reserved slot is added to the set of slots locally reserved for transmission (*MTR* variable). The data source switches to the *"Waiting"* state if its packet queue is empty.

While in the *"Reservation"*, the voice source loses its reserved slot, noted *S*, if another traffic source reserve the slot *S* for reception, i.e., if the data source receives a *ResvCTS* destined to another node and that indicates that slot *S* is being reserved. Another scenario where a data source loses its reservation is when it does not receive an ACK packet after sending a data packet on its reserved slot. In both scenarios, the data source removes the lost slot from the set of slots locally reserved for transmission, i.e., *MTR* variable. If the data source loses all its reservations and has a non-empty packet queue, it returns back to the *"contention"* state and tries to reserve slots.

Note that a data source in the *"Reservation"* state may try to reserve additional slots if the number of reserved slots is lower than the number of queued packets. Thus, in this state, the data source sends its packets on its actually reserved slots, and contends to reserve furthermore slots. When the data source has no packet to transmit in its reserved slot, the slot is removed from the set of slots reserved for transmission, i.e., *MTR* set.

Figure 5-5 shows the state diagram of a node that waits for reservation request for data traffic transmission from its neighbors. First, the data traffic receiver is in the *"Waiting"* state. If a neighbor node sends a reservation request and establishes a successful reservation, the receiver adds the slot to its reception reserved slots set (i.e, *MRR*), and enters the *"Reservation"* state. Otherwise, a reservation failure is considered, and the receiver returns back to the *"Waiting"* state. While in the *"Reservation"* state, the receiver may lose its reservation (and updates its *MRR* variable) if another neighbor node reserves the same slot for transmission. The receiver may also simply release its reserved slot by leaving the ACK mini-slot empty if the sender does not send a data packet on the reserved slot. The receiver returns back to the *"Waiting"* state when it has no reservation, i.e., when $MRR = \Phi$.

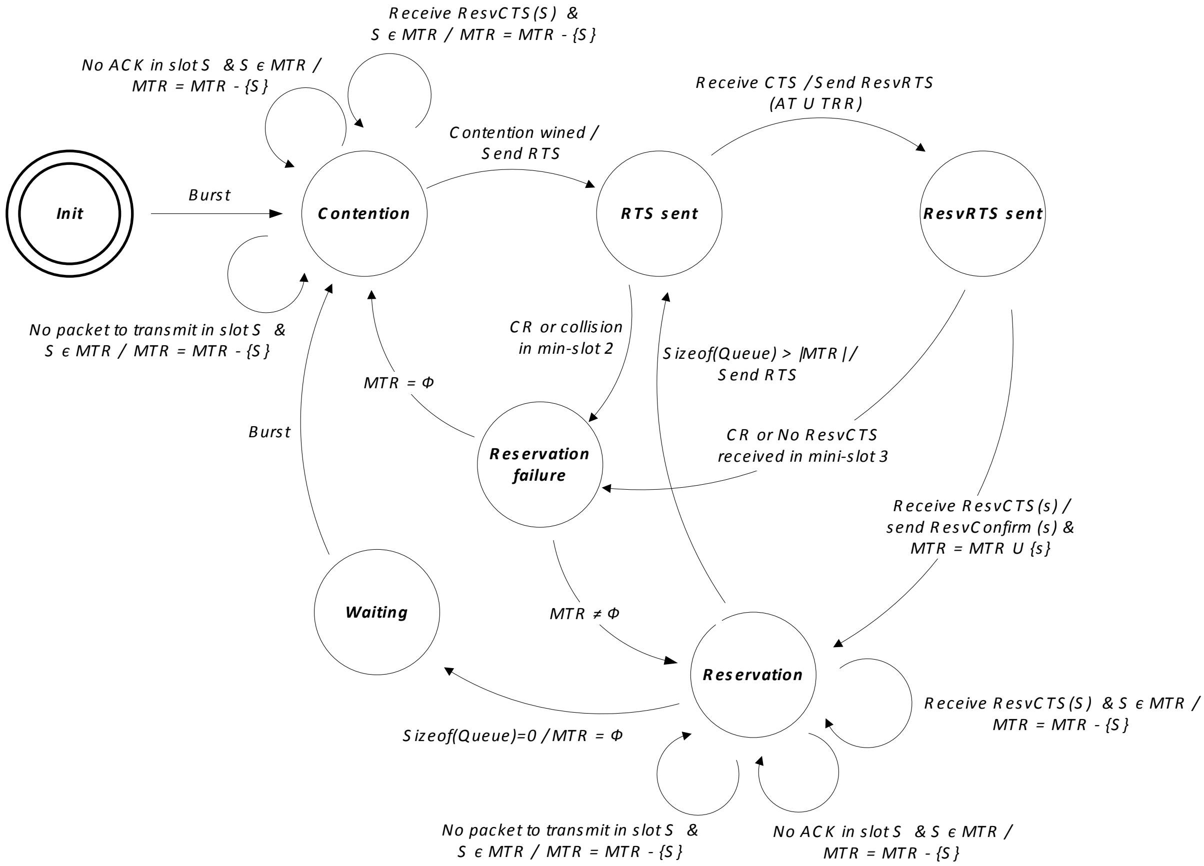

Figure 5-4. State diagram of slot reservation for data sources using the RTR scheme.

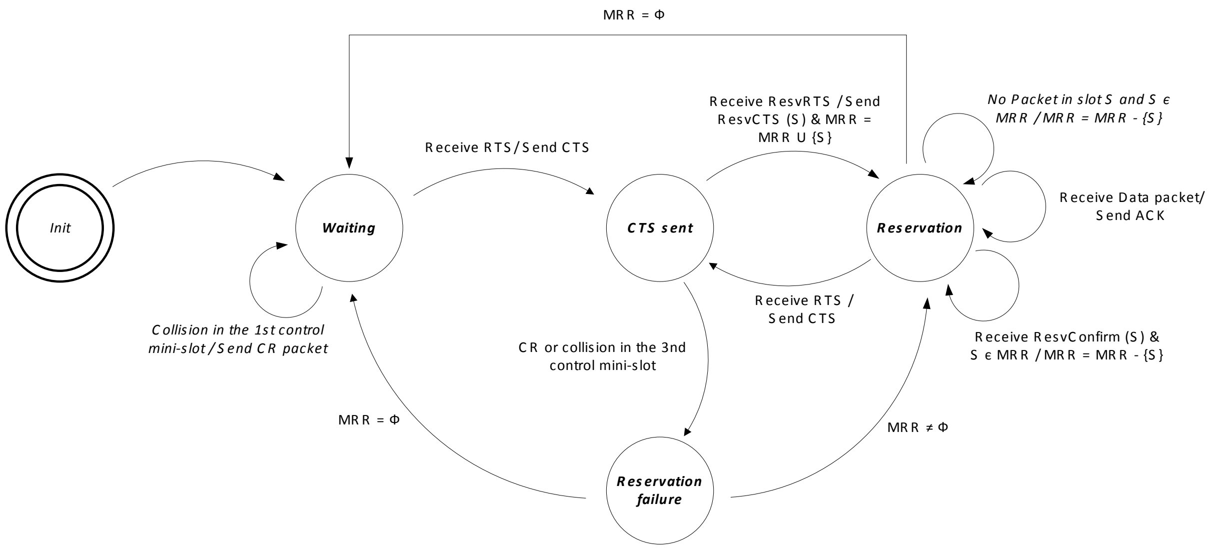

Figure 5-5. State diagram of slot reservation for data traffic reception using the RTR scheme.

5.2.3.1.1 Solution properties

In addition to the properties cited in section 5.2.2.1.1, for a data traffic source the slot state variables with this protocol should meet the following properties:

Property 5-9.

$$\forall i \in N^d, \forall j \in N^v, \forall s, (s \in MTR_i \wedge Neigh1_{ij} = 1) \Rightarrow (s \notin MRR_j \vee s \in MTRR_j)$$

For data traffic source *i*, a slot can be reserved for reception only if it is temporarily transmission released or is not reserved for transmission at all one-hop neighbors of node *i*. Thus,

Property 5-10.

$$\forall i \in N^d, \forall j \in N^v, \forall s, (s \in MRR_i \wedge Neigh1_{ij} = 1) \Rightarrow (s \notin MTR_j \vee s \in MTTR_j)$$

For data traffic source *i*, a slot can be reserved for transmission only if it is temporarily reception released or is not reserved for reception at all one-hop neighbor voice sources of node *i*.

Property 5-11.

$$\forall i, j \in N^d, \forall s, s \in MTR_i \wedge Neigh1_{ij} = 1 \Rightarrow s \notin MRR_j$$

A slot reserved for transmission by data source *i* should not be reserved for reception by any neighbor data sources of node *i*.

Property 5-12.

$$\forall i, j \in N^d, \forall s, s \in MRR_i \wedge Neigh1_{ij} = 1 \Rightarrow s \notin MTR_j$$

A slot reserved for reception by data source *i* should not be reserved for transmission by any neighbor data sources of node *i*.

5.2.3.2 Contend for Each Packet (CEP) scheme

The second access scheme for data traffic sources that we propose is called *Contend for Each Packet (CEP)* scheme. In this scheme, only voice sources are allowed to establish reservations. Data sources send their queued packets using contention on available slots and on temporarily released slots with some permission probability. A data source succeeds in transmitting a data packet on an available slot (or temporarily reception released slot) if no other neighbor of the intended destination contends for transmission of a packet in the same slot.

This scheme has the advantage that it reduces the contention and collision rate of reservation control packets during the *RSF* as only voice sources contend during the *RSF* to transmit their reservation requests. Consequently, time-slot reservation for voice sources is expected to be faster, and packet delay is expected to be lower. However, data traffic delay with this scheme may be higher than with RTR scheme, because unlike RTR scheme where data packets are transmitted without contention, data sources with CEP scheme are required to contend for transmission of each packet. In addition, data traffic sources may suffer excessive dropping rate at high data traffic load due to the high contention rate on available slots.

Finally, the impact of this solution on the performance offered to data traffic depends on the permission probability (i.e., the probability that a node is allowed to send a packet) assigned to data traffic sources.

5.2.3.3 Conclusion

Table 5-6. Data traffic sources access schemes.

Solutions	Description	Advantages	Drawbacks
Solution 1: RTR scheme	Data sources are allowed to reserve slots along with voice sources with priority for voice sources.	Useful for data traffic as data sources are not required to contend for the transmission of each packet.	Increase of contention for voice sources during slot reservation phase as they share the *RSF* with data sources.
Solution 2: CEP scheme	Data sources are allowed to compete during available and temporarily released slots to send their packets.	Low contention during the *RSF* as only voice sources compete for slots reservation.	Low performance expected for data traffic at high data traffic load.

5.2.4 Case study

In this section we illustrate the operations of the protocol through simple examples. We present some scenarios where slot recycling is possible (i.e., situations where data traffic sources use slots not used by voice sources), and other scenarios where slot recycling is not allowed.

In the following scenarios, lines represent point-to-point links, bold arrows represent voice connections with their reserved slot, and dashed arrows represent voice connections with temporarily released slots.

Figure 5-6 shows a scenario of two voice traffic flows *V1* and *V2*. In this scenario, flows *V1* and *V2* are active and cannot reserve the same slot because the receiver of each flow is located in the neighborhood of the sender of the other traffic flow (cf. Properties 5-5 and 5-7)

Figure 5-7 shows a scenario of two voice traffic flows *V1* and *V2*. In this scenario, flow *V2* is active, while *V1* is idle. Flow *V2* cannot reserve slot *S1* reserved by *V1* even if *V1* is idle (cf. Property 5-6 and Property 5-8).

Figure 5-8 shows a scenario of two voice traffic flows *V1* (idle) and *V2* (active), and one data traffic flow *D1*. Since flows V1 and V2 are far away from each other they can reserve the same slot *S1*. Since only voice traffic source *N3* is idle, node *N1* cannot use slot *S1* which is temporarily released by *N3*, because *N6* is still active and uses its reserved slot *S1* for transmission. Thus, with RTR scheme, *D1* can reserve slot *S1* only when slot *S1* is temporarily transmission released by all neighbors of *N2*, i.e, when both of *N3* and *N6* are idle (cf. Property 5-9).

Figure 5-9 shows a scenario of a network with two idle voice traffic flows *V1* and *V2*, and two data traffic flows *D1* and *D3*. Like in the previous scenario both of *V1* and *V2* reserve the same slot *S1* as they don't interfere. Now, both of data flows *D1* and *D2* can reserve slot *S1* without interference during the period of time where slot *S1* is temporarily released for transmission by *N5* and *N7*.

Figure 5-10 shows the same network with two data flows in opposite directions. In this scenario, only one of the two data traffic flows *D1* and *D2* can reserve slot *S1*. Thus, if for example node *N1* reserves slot *S1* with *N2*, *S1* cannot be temporarily reserved for transmission by *N4* because this would lead to a collision at both *N3* and *N2*.

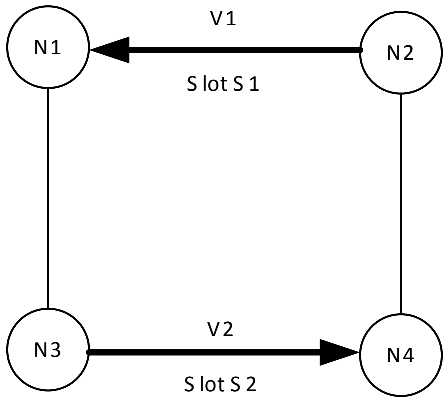

Figure 5-6. Two simultaneous active voice flows that cannot reserve the same slot.

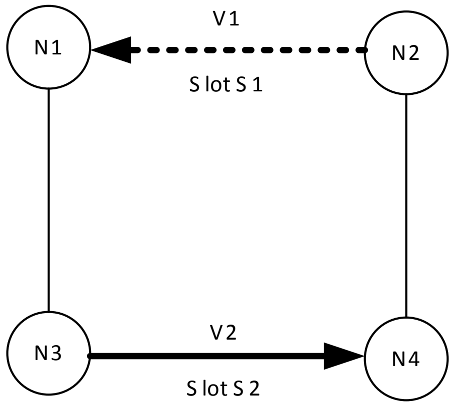

Figure 5-7. Two simultaneous voice flows that cannot reserve the same slot.

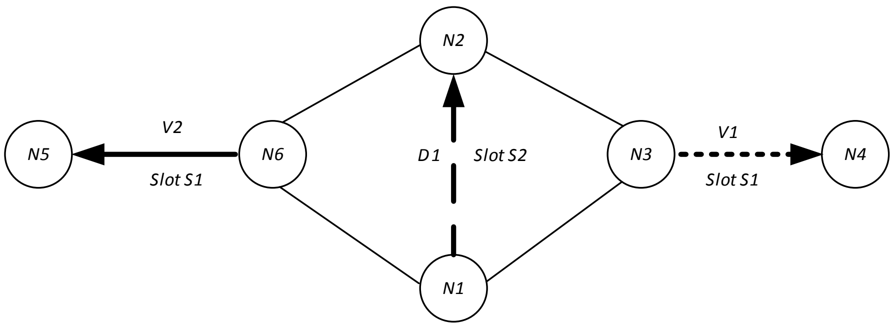

Figure 5-8. Scenario with two voice traffic flows and one data traffic flow.

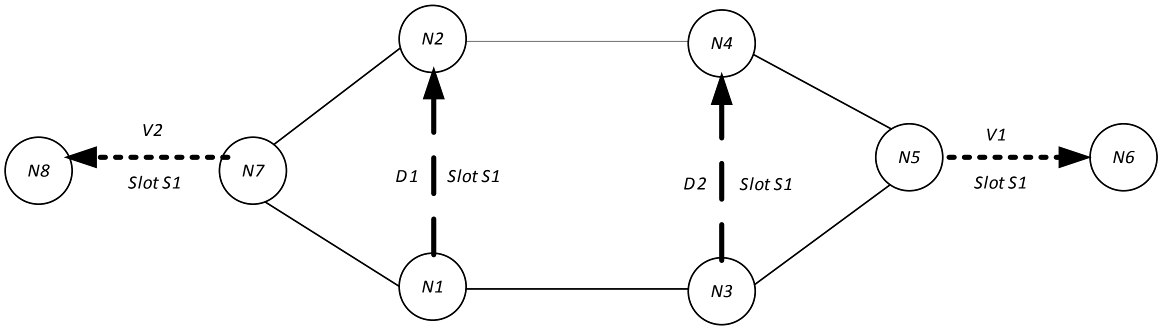

Figure 5-9. A network with two data traffic flows reserving the same slot.

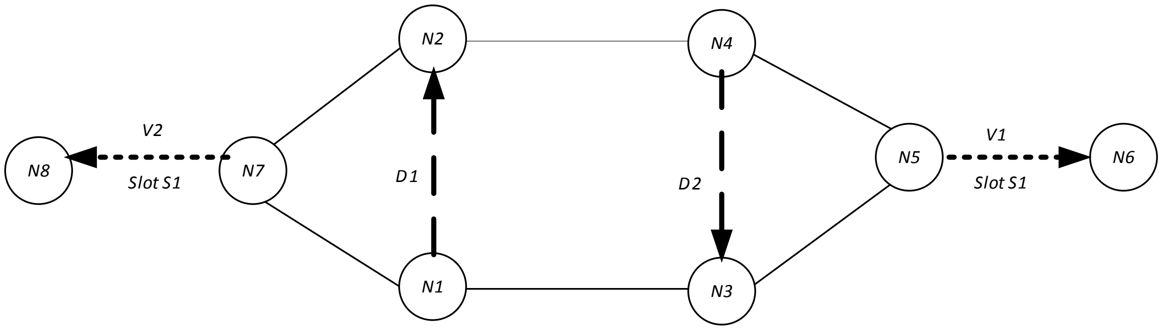

Figure 5-10. A network with two data traffic flows reserving different slots.

5.2.5 Issue 4: Contention Resolution during Reservation Sub-frame

One fundamental part of our reservation scheme is the contention scheme during the *RSF*. This scheme has a great impact on the overall reservation scheme because it controls the fastness of slot reservation (i.e., the reservation establishment delay), which is an important issue for voice traffic sources.

Any Slotted-ALOHA access scheme can be used to resolve contention during the *RSF*. However, the assumption that a voice packet should be transmitted before the arrival of the next one (when a voice source switches to the activity period), leads to the need of a contention resolution scheme that provides fast reservations in order to reduce voice packet queuing delay, and consequently their dropping rate. In this section we propose two access schemes for contention resolution.

5.2.5.1 Solution 1: Static priority contention scheme

In this scheme, all traffic sources send their reservation requests during the *R CRS*s (cf. Figure 3-5) with the same permission probability. If data sources are allowed to reserve slots along with voice sources (the case of RTR scheme), voice sources are given higher priority to send their requests than Data sources. Thus, voice sources send their reservation requests with permission probability p^v, while data sources transmit their requests with priority p^d, such that $p^d<p^v$. The advantage of this contention resolution scheme is its simplicity. However, its performance depends strongly on the values taken by p^d and p^v.

A choice of permission probabilities while suitable for a particular traffic load, it may be unsuitable in other traffic load conditions. On one hand, low values of p^d and p^v are useful in high traffic load as they allow to reduce the contention rate during the reservation sub-frame. However, a choice of low permission probabilities leads to unnecessary additional access delay at low traffic load, because nodes are refrained from sending their reservation requests even if only a few nodes are contending. On the other hand, high values of permission probabilities may be useful at low traffic load, but may lead to excessive collisions of reservation control packets in high traffic load conditions, and consequently to an increase of reservation establishment delay.

5.2.5.2 Solution 2: Dynamic priority contention scheme

In this scheme, the permission probabilities of both voice sources and data sources are adapted to traffic load conditions and collision rate in the *RSF*. We adapt the *Binary Feedback* collision resolution algorithm of Mikhailov [MIK 88] which was proposed to stabilize the Slotted-ALOHA scheme in wireless cellular networks. The algorithm defines a recursive function *S(t)* and updates it slot by slot according to the channel state. *S(t)* is given by:

$$S(t+1) = \begin{cases} \max\{1, S(t) - e + 1\}, & if\ \ feedback(t) = E \\ S(t) + 1, & if\ \ feedback(t) = NE \end{cases} \tag{5-1}$$

Where e is the base of the natural logarithm, $t \in \mathbb{N}^*$, $S(1) = 1$, and E and NE are feedbacks sent by the base station at the end of each slot to indicate that the slot was empty (E) or nonempty (NE). A slot is in the E state if nodes did not transmit packets during this slot, and is NE otherwise. A node which has a packet for transmission in slot t transmits the packet with a permission probability $p(t)=1/s(t)$.

We propose to adapt this algorithm to be used for contention resolution during the *RSF* of our protocol. Every node monitors the channel during the control mini-slots of each *CRS* in order to detect transmissions of its neighbors. Following the status of control mini-slots, we distinguish the three following cases:

Case 1: Idle channel is detected in the 1st and 2nd control mini-slots. The node supposes that there is no neighbor one-hop or two-hops away contenting in the current *CRS*.

Case 2: Reservation failure is detected in the current *CRS*. The failure is detected when collision is detected in the 1st or 2nd control mini-slots, or when a collision is detected or a Collision Report is received in the 2nd mini-slot.

Case 3: Successful reservation is established by a one-hop or a two-hop neighbor. The successful reservation in the current *CRS* is detected when a *ResvCTS* or *ResvConfirm* is received indicating that one or two-hop neighbor is successfully establishing a reservation.

In case 1, a low level of contention is assumed, and the node increases its permission probability accordingly. In case 2, a high level of contention is assumed and the permission probability should be decreased in order to reduce the probability of reservation failure. In case 3, the node maintains the permission probability it used in the previous *CRS*. The new *S* function is expressed as follows:

$$S(t+1)=\begin{cases}\max\{1, S(t)-bonus\}, & if\ Case1\\ S(t)+penality, & if\ Case2\\ S(t), & if\ Case3\end{cases} \qquad (5\text{-}2)$$

Where

Penality=1 and *bonus=e* for voice sources
Penality=e and *bonus=e-1* for data sources

A node which has a reservation request to transmit in the *CRS t* calculates *S(t)* based on the estimate of *S(t-1)* and transmits its reservation request with a permission probability *p(t)=1/S(t)*.

5.2.5.3 Conclusion

Table 5-7. Solutions for contention resolution during the *RSF*.

Solutions	Description	Advantages	Drawbacks
Solution 1	Reserving nodes send their reservation requests during the *RSF* with a static permission probability.	Simplicity.	A choice of the permission probability may not be suitable for all traffic load conditions.
Solution 2	The permission probability of voice and data sources is adapted to the traffic load.	No explicit choice of the permission probability.	Complexity.

The performance of these two contention schemes in terms of the reservation establishment delay depends on the number of *CRS*s composing the *RSF*. Thus, the reservation establishment delay is expected to decrease with the increase of the number of *CRS*s per super-frame. However, the dynamic priority contention scheme is expected to give low reservation establishment delay in comparison with the static priority scheme.

5.3 Conclusion

In this chapter, we presented a solution for bandwidth reservation for voice traffic support over wireless ad-hoc networks. In addition to voice traffic, the proposed solution provides mechanisms to allow data traffic support.

The solution consists in a reservation protocol that controls the medium access of both voice and data sources. We extended the reservation protocol proposed in chapter 3 in order to take into account the characteristics of voice traffic. Indeed, for the design of this protocol, we considered three major issues. The first one is the choice of the data-slot length that provides the best performance. We compared and analyzed three solutions for this issue: long slot length, small slot length with fragmentation, and small slot length without fragmentation.

The second issue is contention resolution during the *RSF*. Contention during the *RSF* has a significant impact on the delay of slot reservation, and consequently on packet transmission delay for voice traffic sources. Thus, in order to provide better performance for voice traffic, we need a robust contention scheme that gives fast reservation establishment even at high traffic load. We proposed two contention resolution schemes that control the permission probability of voice and data sources in sending their reservation requests. In the first scheme, that we called static priority contention scheme, voice and data sources use fixed permission probabilities with higher priority for voice traffic sources. In the second scheme, that we called the dynamic priority contention scheme, the permission probability of voice and data sources is adapted to the traffic load in the network.

The third issue is the way data and voice sources send their packets during the data sub-frame. For voice traffic sources, we showed that reserving bandwidth for voice sources for all the duration of the connection results in a significant waste of bandwidth because voice traffic sources are not active for all the duration of the connection. We proposed to make voice sources release temporarily their reserved slots when they go to the sleep mode, and restoring them when they wake-up, so that data sources can use them for transmission. Data sources can use these temporarily released slots through either contention (in the Contend for Each Packet scheme), or through reserving these slots temporarily (in the Reserve Temporarily Released slots scheme).

In the next chapter, we analyze the performance of this protocol with regard to voice and data traffic through a stochastic model and through simulation.

Chapter 6: Stochastic modeling and performance evaluation of ARPV protocol

6.1 Introduction

In this chapter, we analyze the performance of the reservation protocol proposed in chapter 5 with regard to voice traffic. First, we propose a stochastic model of the protocol and compare its results with those from simulation. This stochastic model is used to analyze and compare the two slot allocation schemes, i.e., the RTR and CEP schemes, and the two contention schemes (i.e., the static and dynamic priority contention schemes). Furthermore, the impact of the slot length on the performance of the protocol is evaluated.

Then, we discuss the obtained results and draw conclusions about the best protocol parameters and the extension mechanisms that enable the protocol to provide the best performance with regard to voice traffic.

6.1.1 Markov chain modeling

In this section, we give a stochastic analysis of our reservation protocol with the different solutions presented in the previous chapter. The first aim of the stochastic analysis is to analyze the impact of the contention schemes during the reservation sub-frame (i.e., static and dynamic priority schemes), and the corresponding permission probabilities on the performance of the protocol. The second aim is to study the impact of the two proposed access schemes for data traffic transmission, i.e., RTR and CEP schemes, on the performance of voice and data traffics.

6.1.1.1 Assumptions

We consider a single-hop network where all nodes are in the transmission range of each other. This network topology assumption is considered for the building of our stochastic model because it simplifies the model. Thus, it is easier to build a stochastic model for a single-hop network than for a multi-hop distributed network. However, as the super-frame is the same and as voice and data traffics pattern are independent of the network topology, we consider that the protocol parameters that give the best performance in a single-hop network (for example the permission probabilities) can be used for a multi-hop network.

We observe that the behavior of the protocol in a single-hop network can be fully described by some state variables of a Markov process, describing the status of voice and data sources. A model is used to derive metrics of interest like voice packet dropping rate, and data traffic delay and throughput. We model the protocol with the static and dynamic priority contention schemes, and RTR and CEP reservation schemes.

6.1.1.2 Network, voice and data source models

The super-frame length, noted *SFL*, is composed of *S* data slots. We consider a single-hop network with N^v voice nodes and N^d data nodes. Each voice source can get a reservation for at most one time-slot. Since a slot is dedicated to one voice source, the number of voice sources in the network is bounded by the number of data slots, i.e., $N^v \le S$.

Data sources are allowed to send more than one packet per super-frame. Thus, with the RTR scheme, data sources can reserve more than one slot per super-frame when several voice sources are silent or when several slots are available.

In a single-hop network with several contending nodes, the successful reservation of a data slot depends only on the state of channel access during the first mini-slot of *CRS*s. If among all contending nodes exactly one node transmits a *ResvRTS* in the first mini-slot of a *CRS*, such a node establishes a successful reservation. If two or more nodes transmit *ResvRTS* packets in the first mini-slot of the same *CRS*, then these transmissions are unsuccessful.

6.1.1.2.1 Voice source model

Each voice source is equipped with a voice activity detector (VAD) and may be, either in *silence* state, in *reservation* state, or in *contention* state. A voice source is in the *silence* state when it has no packet to transmit. In this state, the voice source releases temporarily its reserved slot by leaving it empty.

A transition to *contention* state occurs when a talkspurt is generated and the voice source has not yet restored its reserved slot. Voice source stays in this state and competes to reserve a slot. When it successfully reserves a slot, it enters *reservation* state. Transitions between these states are shown in Figure 6-1, where p_{On} is the probability that a talkspurt is generated in a super-frame, p_{Off} is the probability that a talkspurt ends in a super-frame, p^v is the permission probability of voice sources, and p^v_{Succ} is the probability that the voice source gets a reservation in the current super-frame. We assume the length of talkspurt and silence periods are exponentially distributed with means t_{On} and t_{Off} durations respectively. p_{On} and p_{Off} are expressed as follows [GOD 88]:

$$p_{On} = 1 - \exp(-SFL/t_{On}) \tag{6-1}$$

$$p_{Off} = 1 - \exp(-SFL/t_{Off}) \tag{6-2}$$

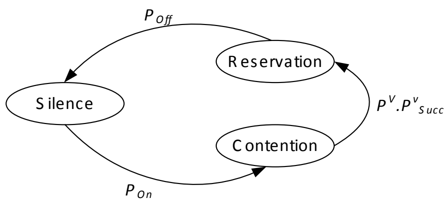

Figure 6-1. State-transition diagram of voice source.

We assume that silent nodes that begin a talkspurt in the middle of the super-frame wait until the beginning of the next super-frame to contend for time-slot reservation. The voice source attempts to reserve a slot at the arrival of the initial packet of a talkspurt. It drops its queued packet if a new voice packet arrives while reservation is still not established. If the source drops a packet of a talkspurt, it continues contention for a slot reservation to send subsequent packets.

6.1.1.2.2 Data source model

We assume data packets at each data source arrive following a Poisson process with equal mean arrival rate of λ data packets/super-frame. The minimum and the maximum number of packets that can be generated by a data source are noted λ_{min} and λ_{max} respectively. Data packets are assumed to arrive in the form of bursts, and arrive at the end of the super-frame. Let $P_i(arrival=x)$ denote the probability that a burst of x packets arrive to the data source i at the end of a super-frame. $P_i(arrival=x)$ is given by:

$$\mathrm{P_i}(\text{arrival} = \mathrm{x}) = e^{-\lambda\frac{\lambda^x}{x!}} \tag{6-3}$$

Note that this probability is the same for all data sources.

Packets that arrive at a data source are stored in a *FIFO* packet queue with a finite capacity *Buffer_size*. When the packet queue is full any arriving packets are dropped. A data source without backlogged packets is in the thinking state waiting for the generation of data packets.

In the RTR scheme, data sources attempt to reserve slots each time they have bursts. At the beginning of each super-frame, each data source checks its packet queue. If the number of backlogged packets is higher than the number of reserved slots, the data source competes during the *RSF* to reserve slots. In order to provide fairness between backlogged data sources and avoid that some data sources with long packet queue reserve high number of slots, each data source is allowed to reserve only one slot during a super-frame. The data source can reserve further slots, by repeating the contention process in subsequent super-frames.

Each data source n is characterized by two state variables: the number of backlogged packets $Q[n]$, and the number of slots reserved for this data source $R^d[n]$. Thus, the number of slots reserved to the whole data sources at any time is represented by a vector state variable $R^d \in \mathrm{Z}^{|N^d|}$, where each entry $R^d[n]$ being the number of slots reserved to data source n and is upper-bounded by the number of data-slots, i.e. $0 \le \mathrm{R^d[n]} \le \mathrm{S}$. The number of backlogged packets for the whole data sources is represented as a vector state variable $Q \in \mathrm{Z}^{|N^d|}$, where each entry $Q[n]$ represents the number of backlogged packets at data source n, with $0 \le Q[n] \le Buffer\_size$.

The set of all possible states for the number of reservations of all data sources is denoted by R^d*_states*, and the set of possible states for the backlogged traffic for all data sources is denoted by *Q_states.*

$Q[n]$ increases when the number of packets arriving to the data source n is higher than the number of slots reserved to it. It decreases when this number is lower than $R^d[n]$. A data source n requests reservation of additional slots if $R^d[n]<Q[n]$. The number of slots reserved to the data source n can increase if there are available slots or if some voice sources are in the silence state. Data source n may

lose a part of its reservations while it has backlogged packets and *Q[n]* decreases if a voice source grabs slots reserved to this data source due to slots non availability.

In the CEP scheme, data sources don't reserve slots, but contend during available slots or temporarily released slots to transmit their queued packets. Successful transmission in a slot does not give to the data source the right to transmit during the same slot in subsequent super-frames. The state variables needed to model this scheme are the number of voice sources in *reservation* state R^v, and the number of backlogged packets by all data sources Q.

6.1.1.3 Markov Chain Model of Voice and Data System

6.1.1.3.1 Steady state distribution of the RTR scheme

Our reservation scheme with the RTR scheme can be fully described by four state variables $\{R^v, R^d, Q, Sil\}$, namely, the number of voice sources in the reservation state $R^v \in Z$, the number of slots reserved to each data source $R^d \in Z^{|N^d|}$, the number of backlogged packets at each data source $Q \in Z^{|N^d|}$, and the number of voice sources in the silence state $0 \le Sil \le N^v$.

The number of voice sources in the contention state C^v is:

$$C^v = |N^v| - R^v - Sil \qquad (6\text{-}4)$$

Let us define the following random variables:

$R^v(t)$: The number of voice sources in the *Reservation* state at the end of the *t-th* super-frame.

$C^v(t)$: The number of voice sources in the *Contention* state at the end of the *t-th* super-frame.

Sil(t): The number of voice sources in the *Silence* state at the end of the *t-th* super-frame.

$R^d(t)$: The state vector describing the number of slots reserved to each data source at the end of the *t-th* super-frame.

Q(t): The state vector describing the number of backlogged packets at each data source at the end of the *t-th* super-frame. The length of vectors R^d and Q is equal to the number of data sources, i.e. $R^d, Q \in Z^{|N^d|}$.

Since the state of the system in the current super-frame depends only on its state in the previous super-frame, the evolvement of the system can be modeled as a 4-D Markov process, with the system state representing the four variables $\{R^v, R^d, Q, Sil\}$ at the end of each super-frame. Since the system is irreducible and has a finite number of states, the stationary distribution of the system is supposed to exit. Let $\Pi^{(rv,rd,q,sil)}$ denote the stationary distribution of this system.

$$\prod^{(rv,rd,q,sil)} = \left\{\pi(r_i^v, r_i^d, q_i, sil_i)\right\} = \left\{P(R^v = r_i^v, R^d = r_i^d, Q = q_i, Sil = sil_i)\right\} \qquad (6\text{-}5)$$

Where $r_i^v \in [0,|N^v|]$, $r_i^d \in R^d\_states$, $q \in Q\_states$, $sil_i \in [0,|N^v|]$. Solving this 4-D Markov process is computationally complex. Therefore, since the number of voice sources in the silence/talking state is independent of the number of reserved slots (by voice and data sources), and depends only on the talkspurt/silence pattern of voice sources, solving this 4-D Markov chain can be reduced to solving two sub-processes: the silence/talking sub-process Π^{Sil} with a state variable *Sil* describing the transition of voice sources between silence and talking states, and the reservation sub-process which describes the contention/reservation state of voice/data sources with three state variables $\{R^v, R^d, Q\}$. The evolvement of the reservation sub-process depends on the number of silent nodes in the system (i.e. *Sil*) which is obtained by solving the Π^{Sil} system. Let us denote this system by $\Pi^{(rv,\, rd,\, q|\, sil)}$. We have:

$$\prod^{Sil} = \{\pi(sil_i)\} = \{P(Sil = sil_i)\} \tag{6-6}$$

$$\prod^{(rv,rd,q|sil)} = \left\{\pi(r_i^v, r_i^d, q_i \middle| sil_i)\right\} = \left\{P(R^v = r_i^v, R^d = r_i^d, Q = q_i \middle| Sil = sil_i)\right\} \tag{6-7}$$

The initial 4-D system can be easily solved after the steady states of both sub-systems Π^{Sil} and $\Pi^{(rv,\, rd,\, q|sil)}$ are found. Let us first compute the transition probability matrix P^{Sil} of the silence/talkspurt system Π^{Sil}:

$$\mathrm{P}^{\mathrm{Sil}} = \{\mathrm{P(Sil(t+1) = j\,|\,Sil(t) = i)}\} \tag{6-8}$$

Where $i, j = 0, \ldots, |N^v|$, and $Sil(t)$ represents the number of voice sources in the silence state at the end of the *t-th* super-frame. Transition probability of the number of voice nodes in the silence state from the super-frame *t* to super-frame *t+1* is given by:

$$P(Sil(t+1) = j \middle| Sil(t) = i) = \sum_{k=\max(0,i-j)}^{\min(i,N^v-j)} B(i,k,p_{On}).B(N^v - i, j-(i-k), p_{Off}) \tag{6-9}$$

Where $B(n,m,p)$ is the Binomial distribution given by

$$B(n,m,p) = \binom{n}{m} p^m (1-p)^{n-m} \tag{6-10}$$

Based on transition probabilities, we can find the steady state distribution of the silence/talkspurt system by solving the following equation system:

$$\prod^{Sil} = \prod^{Sil}.P^{Sil} \tag{6-11}$$

and

$$\sum_{sil_i} \pi(sil_i) = 1 \tag{6-12}$$

Now, we compute the steady state distribution of the reservation process $\Pi^{(rv,\, rd,\, q|sil)}$ given the assumption that the number of voice sources in the silence state (*Sil)* is known. In order to compute the transition probability matrix, we define the following probability functions:

- $\varphi_v(v, d, p^v, p^d)$: the probability that with *v* voice sources in contention state and *d* data sources in contention state, there is a voice source which succeeds in establishing a reservation in the current *CRS* (i.e. only one voice source transmits its *ResvRTS* among the *v+d* contending nodes). p^v and p^d are the permission probabilities for voice and data sources respectively. $\varphi_v(v,d, p^v, p^d)$ is given by:

$$\varphi_v(v, d, p^v, p^d) = B(v, 1, p^v) \cdot B(d, 0, p^d) \tag{6-13}$$

- $\varphi_d(v, d, p^v, p^d)$: the probability that with *v* voice sources in contention state and *d* Data sources in contention state, there is a data source which succeed establishing a reservation in the current CRS. $\varphi_d(v,d, p^v, p^d)$ is given by

$$\varphi_d(v, d, p^v, p^d) = B(d, 1, p^d) \cdot B(v, 0, p^v) \tag{6-14}$$

- $\Psi_{staticP}(v, d, c, v1, d1)$: the probability that with the static priority contention scheme, among *v* voice nodes and *d* data nodes in contention state, there are *v1* voice successful reservations and *d1* data successful reservations in *c CRSs*.

$$\begin{aligned}\psi_{staticP}(v,d,c,v1,d1) = {} & [1-\varphi_v(v,d,p^v,p^d)-\varphi_d(v,d,p^v,p^d)]\cdot\psi_{staticP}(v,d,c-1,v1,d1) + \\ & \varphi_v(v,d,p^v,p^d)\cdot\psi_{staticP}(v-1,d,c-1,v1-1,d1) + \varphi_d(v,d,p^v,p^d)\cdot\psi_{staticP}(v,d-1,c-1,v1,d1-1)\end{aligned} \tag{6-15}$$

Where the ending condition of this recursive function is:

$$\psi_{staticP}(v,d,c,v1,d1) = \begin{cases}[1-\varphi_v(v,d,p^v,p^d)-\varphi_d(v,d,p^v,p^d)]^c & if\ v1 = d1 = 0 \\ 0 & if\ c < v1 + d1\, or\ v1 < 0\, or\ d1 < 0\end{cases} \tag{6-16}$$

- $\Psi_{dynamicP}(v, d, c, v1, d1, S^v, S^d)$: the probability that with the dynamic priority contention scheme, among *v* voice nodes and *d* data nodes in the contention state, there are *v1* voice successful reservations and *d1* data successful reservations in *c CRSs*. S^v and S^d are the values given by the *S(t)* function given in formula (5-2).

$$\begin{aligned}\psi_{dynamicP}(v,d,c,v1,d1,S^v,S^d) = {} & [1-\varphi_v(v,d,1/S^v,1/S^d)-\varphi_d(v,d,1/S^v,1/S^d)]\cdot[\psi_{dynamicP}(v,d, \\ & c-1,v1,d1,S^v+1,S^d+e) + \psi_{dynamicP}(v,d,c-1,v1,d1,\max(1,S^v-e),\max(1,S^d-(e-1)))] + \\ & \varphi_v(v,d,1/S^v,1/S^d)\cdot\psi_{dynamicP}(v-1,d,c-1,v1-1,d1,S^v,S^d) + \\ & \varphi_d(v,d,1/S^v,1/S^d)\cdot\psi_{dynamicP}(v,d-1,c-1,v1,d1-1,S^v,S^d)\end{aligned} \tag{6-17}$$

The ending condition of this recursive function is:

$$\psi_{dynamicP}(v,d,c,v1,d1,S^v,S^d) = \begin{cases}1-\varphi_v(v,d,1/S^v,1/S^d)-\varphi_d(v,d,1/S^v,1/S^d) & if\ v1 = d1 = 0\, and\ c = 1 \\ 0 & if\ c < v1 + d1\, or\ v1 < 0\, or\ d1 < 0\end{cases} \tag{6-18}$$

- $\Re(Vc, Dc, Vr, Dr, l, p)$: the probability that among *Vc* voice sources and *Dc* data sources in contention state, and with *Vr* voice source and *Dr* data sources in reservation state, *l* voice sources and *p* data sources obtain a reservation in the current super-frame. $\Re(Vc, Dc, Vr, Dr, l, p)$ for the static priority scheme is given by:

$$\Re(Vc, Dc, Vr, Dr, l, p) = \begin{cases} 0, & p > S - Vr - Dr \; or \; l > S - i \\ \psi_{staticP}(Vc, Dc, R, l, 0), & p = 0 \; and \; l \le S - Vr \\ \psi_{staticP}(Vc, Dc, R, l, p), & p > 0 \; and \; l + p \le S - Vr - Dr \end{cases} \quad (6\text{-}19)$$

$\Re(Vc, Dc, Vr, Dr, l, p)$ for the dynamic priority scheme is given by:

$$\Re(Vc, Dc, Vr, Dr, l, p) = \begin{cases} 0, & p > S - Vr - Dr \; or \; l > S - i \\ \psi_{dynamicP}(Vc, Dc, R, l, 0, 1, 1), & p = 0 \; and \; l \le S - Vr \\ \psi_{dynamicP}(Vc, Dc, R, l, p, 1, 1), & p > 0 \; and \; l + p \le S - Vr - Dr \end{cases} \quad (6\text{-}20)$$

Transition probability of the reservation sub-process $\Pi^{(rv,\, rd,\, q|sil)}$ depending on the number of silent nodes (*Sil=sil*) is given by formula (6-24). In formula (6-24) *unsatisfied_nodes(t)* denotes the set of data sources for which the number of backlogged packets at the end of the *t-th* super-frame is higher than the number of reserved slots. It is calculated as follows:

$$unsatisfied\_nodes(t) = \left\{ i \in N^d \,\middle|\, Q[i](t) > R^d[i](t) \right\} \quad (6\text{-}21)$$

Resv^d(t) designates the number of slots reserved to all data sources at the end of the *t-th* super-frame, and is calculated as follows:

$$\mathrm{Re}sv^d(t) = \sum_{n=1}^{|N^d|} R^d[n](t) \quad (6\text{-}22)$$

Success^d(t+1) is the set of data sources who obtain reservations successfully during the *(t+1)-th* super-frame. It is calculated as follows:

$$success^d(t+1) = \left\{ i \in N^d \,\middle|\, R^d[i](t+1) > R^d[i](t) \right\} \quad (6\text{-}23)$$

$$P(R^v(t+1)=i2, R^d(t+1)=j2, Q(t+1)=b2 \,\big|\, R^v(t)=i1, R^d(t)=j1, Q(t)=b1) =$$

$$\begin{cases} 0, & \text{if } \exists n \in N^d \,\big|\, R^d[n](t+1) > R^d[n](t)+1 \\ 0, & \text{if } \exists n \in N^d \,\big|\, Q[n](t) < R^d[n](t) \text{ and } R^d[n](t+1) > R[n](t) \\ 0, & \text{if } \exists n \in N^d \,\big|\, Q[n](t) \geq R^d[n](t) \text{ and } R^d[n](t+1) < R^d[n](t) \text{ and } i2 + \mathrm{Re}\,sv^d(t) < S \\ \left[\displaystyle\sum_{k=\max(0,u-i)}^{\min(N^v-sil-i,S-i)} \Re(N^v - sil - i, |unsatisfied\_nodes(t)|, i1, \mathrm{Re}\,sv^d(t), k, |success^d(t+1)|) \cdot B(i+k, u, 1-p_{Off})\right] \cdot & \\ \dbinom{|unsatisfied\_nodes(t)|}{|success^d(t+1)|} \cdot \displaystyle\prod_{n=1}^{|N^d|} P_n(arrival = Q[n](t+1) + R^d[n](t+1) - Q[n](t)), & otherwise \end{cases}$$

(6-24)

The first case in formula (6-24) corresponds to the situation where there exists a data source that reserves more than one slot during a super-frame. The probability of this transition is zero because we assumed that a backlogged data source is allowed to increase its reservations only by one slot during a super-frame even if this source has more than one backlogged packet. The second case is to avoid transitions where some data sources reserve more slots than what they need. The third case corresponds to the situation where the number of slots reserved to a backlogged data source with a backlog exceeding the number of reserved slots decreases while there are available slots. The number of slots reserved to a data source decreases only if the backlog is lower than the actual reservations (i.e. some reserved slots must be released), or if the number of voice sources who establish reservations increases (these voice sources grab slots reserved to the data source) and that there is no available slots. The last case corresponds to the situation where among N^v-*sil*-*i* voice sources in the contention state and among the data sources which have queue length higher the number of reserved slots, *k* voice source obtains a successful reservation and $|success^d(t+1)|$ data source obtains a successful reservation in the *(t+1)-th* super-frame.

With these one-step transition probabilities, we can compute the transition probability matrix $P^{(rv,\, rd,\, q|sil)}$ and find the steady state distribution of the reservation sub-system $\Pi^{(rv,\, rd,\, q|sil)}$ depending on the number of voice sources in silence state by solving the following equation systems:

$$\Pi^{(rv,rd,q|sil)} = \Pi^{(rv,rd,q|sil)}.P^{(rv,rd,q|sil)} \quad and \quad \sum_{r_i^v}\sum_{r_i^d} \pi(r_i^v, r_i^d, q|sil) = 1 \qquad (6\text{-}25)$$

6.1.1.3.2 Steady state distribution of the CEP scheme

In this scheme, data sources do not reserve slots, but compete for the transmission of each data packet. To model this scheme, we need to know the number of backlogged packets at each data source, and the number of voice sources in *reservation* state depending on the number of voice sources in the silence state. Let $\Pi^{(rv,\, q|sil)}$ denote the stochastic process describing the evolvement of the number of voice sources with reservation and the number of backlogged data packets.

$$\prod^{(rv,q|sil)} = \left\{\pi(r_i^v, q_i | sil_i)\right\} = \left\{P(R^v = r_i^v, Q = q_i | Sil = sil_i)\right\} \qquad (6\text{-}26)$$

Let us compute the steady state distribution of this system given the assumption that the number of voice sources in the silence state is known. The transition probabilities of this system are as follows:

$$P(R^v(t+1)=i2, Q(t+1)=q2 \,|\, R^v(t)=i1, Q(t)=q1)=$$

$$\begin{cases} 0, & if \quad \sum_{i=1}^{|N^d|} \max(q1[i]-q2[i],0) > S-i1 \\ \left[\sum_{k=\max(0,i2-i1)}^{\min(N^v-sil-i1,S-i1)} \Re(N^v - sil - i1, 0, i1, 0, k, 0) \,.\, B(i1+k, i2, 1-p_{Off}) \right]. & \\ \sum_{a\in Arrival\_states} \Phi(|back\log ged\_nodes(t)|, q1, S-\max(i1,i2), reqSlots(t+1), a).\prod_{n=1}^{|N^d|} P_n(arrival = a[i]), & otherwie \end{cases}$$

(6-27)

with

$$\Phi(contenders, back\log, C, \mathrm{Re}\,q, arrival) = \begin{cases} 0, & if\ \exists i \in N^d \,|\, q2[i] > q1[i] + arrival[i]\ or\ q2[i] < arrival[i] \\ \psi_{CEP}(contenders, back\log, C, \mathrm{Re}\,q), & otherwise \end{cases} \quad (6\text{-}28)$$

$$\psi_{CEP}(contenders, back\log, C, \mathrm{Re}\,q) = \sum_{i\in N^d / back\log[i]>0\ and\ \mathrm{Re}\,q[i]>0} \zeta(contenders, p^d).\frac{1}{contenders}.$$
$$\psi_{CEP}(contenders-1, dec(back\log, i), C-1, dec(\mathrm{Re}\,q, i)) + (1-\zeta(contenders, p^d)).$$
$$\psi_{CEP}(contenders, back\log, i), C-1, \mathrm{Re}\,q) \quad (6\text{-}29)$$

The ending condition of this recursive function is:

$$\psi_{CEP}(contenders, back\log, C, \mathrm{Re}\,q) = \begin{cases} (1-\zeta(contenders, p^d))^c, & if \sum_{i\in N^d} \mathrm{Re}\,q[i] = 0 \\ 0, & if \sum_{i\in N^d} \mathrm{Re}\,q[i] > C \end{cases} \quad (6\text{-}30)$$

Where

- *backlogged_nodes(t)* is the set of data sources with backlogged traffic at the end of the *t-th* super-frame, and is calculated as follows:

$$back\log ged\_nodes(t) = \{ i \in N^d / Q(t)[i] \geq 1 \} \quad (6\text{-}31)$$

- $reqSlots(t+1) \in \mathrm{Z}^{|N^d|}$ is the vector giving the number of slots required by each data source at the beginning of the *(t+1)-th* super-frame. *reqSlots(t+1)[i]* designates the number of slots which should be allocated to data source *i* so that its backlogged traffic changes from *q1[i]* to *q2[i]*. *reqSlots* is calculated as follows:

$$reqSlots(t+1)[i] = q1[i] + a[i] - q2[i] \quad (6\text{-}32)$$

- Ψ_{CEP}*(contenders, backlog[], C, Req[])* is a recursive function which permits to calculate the probability that each data source *i* transmits *Req[i]* packet successfully in a super-frame. *Contenders* and *backlog[]* are the number of data sources in contention and the number of backlogged packets of each data source respectively. *C* is the number of available slots, i.e. slots not reserved by voice sources or temporarily released. In formula (6-29), the *dec(V, i)* is a function that returns a vector similar to *V* with the *i–th* entry decremented by one.

- ξ *(c, p)* is the probability that with *c* data sources in contention, one data source transmits a data packet in the current data slot. *p* is the permission probability of data sources during available slots. ξ *(c, p)* is given by:

$$\xi(c,p) = B(1,1,p)\cdot B(c-1,0,p) \tag{6-33}$$

The first case in formula (6-27) ensures that the decrease of backlogged traffic at all nodes does not exceed the number of slots which are not reserved by voice sources.

6.1.1.4 Voice traffic performance

We use the defined Markov system to derive some performance measurement metrics. The main performance measure for voice traffic is packet dropping rate. Voice packet dropping are mainly due to excessive packet queueing. The average voice packet dropping rate is calculated as [RS 92] using the following formula:

$$P_{drop}{}^{v} = \frac{E(C^v) - E(S^v)}{E(C^v) + E(R^v)} \tag{6-34}$$

Where $E(C^v)$ is the average number of voice sources in the contention state, $E(R^v)$ is the average number of voice sources with reservation, and $E(S^v)$ is the average number of successful voice reservations per super-frame.

$$E_{RTR}(C^v) = \sum_{r^v=0}^{N^v} \sum_{sil=0}^{N^v-r^v} \sum_{r^d \in R^d\_states} \sum_{q \in Q\_states} (N^v - r^v - sil).\pi(r^v, r^d, q, sil) \tag{6-35}$$

$$E_{CEP}(C^v) = \sum_{r^v=0}^{N^v} \sum_{sil=0}^{N^v-r^v} \sum_{q \in Q\_states} (N^v - r^v - sil)\pi(r^v, q, sil) \tag{6-36}$$

$$E_{RTR}(R^v) = \sum_{r^v=0}^{N^v} \sum_{sil=0}^{N^v-r^v} \sum_{r^d \in R^d\_states} \sum_{q \in Q\_states} r^v \pi(r^v, r^d, q, sil) \tag{6-37}$$

$$E_{CEP}(R^v) = \sum_{r^v=0}^{N^v} \sum_{sil=0}^{N^v-r^v} \sum_{q \in Q\_states} r^v \pi(r^v, q, sil) \tag{6-38}$$

$$E_{RTR}(S^v) = \sum_{r^v=0}^{N^v} \sum_{sil=0}^{N^v-r^v} \sum_{r^d \in R^d\_states} \sum_{q \in Q\_states} \left(\sum_{v=0}^{N^v-r^v-sil} \sum_{d=0}^{N^d-r^d} v \cdot \Re(r^v, r^d, N^v - r^v - sil, N^d - r^d, v, d) \right). \pi(r^v, r^d, q, sil) \quad (6\text{-}39)$$

$$E_{CEP}(S^v) = \sum_{r^v=0}^{N^v} \sum_{sil=0}^{N^v-r^v} \sum_{q \in Q\_states} \left(\sum_{v=0}^{N^v-r^v-sil} l . \Re(r^v, 0, N^v - r^v - sil, 0, v, 0) \right) . \pi(r^v, q, sil) \quad (6\text{-}40)$$

6.1.1.5 Data traffic performance

While satisfying a low dropping rate for voice traffic, the system should offer some acceptable service to data sources. Let P_{drop}^d denote the data packet dropping rate. P_{drop}^d is given by:

$$P_{drop}^d = \frac{E(D^d)}{E(A)} \quad (6\text{-}41)$$

Where $E(A)$ is the average number of arrived data packets at all data sources during a super-frame. As the average arrival rate of data packets at a data source is λ, and is the same for all data sources, we have:

$$E(A) = \lambda . \left| N^d \right| \quad (6\text{-}42)$$

$E(D^d)$ is the average number of data packets dropped at the end of a super-frame because the packets queue is full, and is given by:

$$E_{RTR}(D^d) = \sum_{r^v=0}^{N^v} \sum_{sil=0}^{N^v-r^v} \sum_{r^d \in R^d\_states} \sum_{q \in Q\_states} \left(\sum_{i=1}^{|N^d|} \sum_{a=\lambda_{\min}}^{\lambda_{\max}} P_i(arrival = a) . \max(0, q[i] + a - Buffer\_size) \right) . \pi(r^v, r^d, q, sil) \quad (6\text{-}43)$$

$$E_{CEP}(D^d) = \sum_{r^v=0}^{N^v} \sum_{sil=0}^{N^v-r^v} \sum_{q \in Q\_states} \left(\sum_{i=1}^{|N^d|} \sum_{a=\lambda_{\min}}^{\lambda_{\max}} P_i(arrival = a) . \max(0, q[i] + a[i] - Buffer\_size) \right) . \pi(r^v, q, sil) \quad (6\text{-}44)$$

We compute the average delay using the Little's formula: the average time that a packet stays in a queuing system is computed as the ratio of the average number of packets waiting for transmission to the average arrival rate of packets. The average delay for data traffic in term of number of super-frames is given by:

$$Delay^d = \frac{E(Q)}{E(A)} \quad (6\text{-}45)$$

Where *E(Q)* is the average queued packets at all data sources, and is given by:

$$E_{RTR}(Q) = \sum_{r^v=0}^{N^v} \sum_{sil=0}^{N^v-r^v} \sum_{r^d \in R^d\_states} \sum_{q \in Q\_states} \left(\sum_{i=1}^{|N^d|} q[i] \right) \pi(r^v, r^d, q, sil) \qquad (6\text{-}46)$$

$$E_{CEP}(Q) = \sum_{r^v=0}^{N^v} \sum_{sil=0}^{N^v-r^v} \sum_{q \in Q\_states} \left(\sum_{i=1}^{|N^d|} q[i] \right) . \pi(r^v, q, sil) \qquad (6\text{-}47)$$

6.2 Performance evaluation

In this section, we analyze the performance of our protocol in different scenarios. First, we compare its behavior with the RTR and CEP reservation schemes. In addition, the impact of both contention resolution schemes on the performances of the RTR and CEP schemes is evaluated. We consider a single-hop network where analytical results are compared with those of simulation. Afterwards, we evaluate the performance of the protocol in a multi-hop network where the impact of the different slot length choices is analyzed. For simulation, we use the NS-2 simulator [NS2].

6.2.1 Stochastic vs. simulation results in a single-hop network

We present stochastic results of the RTR and CEP reservation schemes with dynamic and static priority contention schemes, using the performance evaluation model described in section 6.1.1 in a single-hop network. Analytical results are compared to those from simulation under default system parameters specified in Table 6-1.

This section allows us to compare the performance obtained with the static and dynamic priority schemes, and determine the permission probabilities of the static priority scheme that give the best performance.

6.2.1.1 Simulation model

We consider a single-hop network composed of a number of voice and data sources. We consider a wireless channel of 2 Mbps bit rate. We do not consider the slot length issue in this section. For instance, we consider the first slot length choice solution, i.e., the slot length is set to the packet length of the G.711 codec. Thus, voice sources are considered equipped with G.711 encoder that generates packets of 160 bytes payload at 64 kbps rate. The talkspurt and silence durations are 1.35s and 1s respectively [GOD 88, KPP 05].

The super-frame is considered composed of 10 *CRS*, and 12 data slots. According to the protocol description, the maximum number of voice sources which can be admitted into the network is 12 voice sources. Each Data slot consists of the transmission time of one voice packet with the different layers overhead. With 2 Mbps channel bit-rate, the Data slot length is 1.024 ms. Each voice source is required to reserve a slot per super-frame in order to send its voice packets.

For data sources, the mean packet generation rate, λ, is set to one packet per super-frame. The minimum and maximum packet generation rate are λ_{Min}=0 and λ_{Max}=3 respectively. General simulation and protocol parameters considered in this section are shown in Table 6-1.

Table 6-1. Simulation and default system parameters.

Parameter	Value
Channel bit rate (Mbps)	2
RTP+UDP+IP header (bytes)	12+8+20
MAC header (bytes)	18
PHY layer overhead (PLCP header + preamble) (bits)	48+56
Data slot payload size (bytes)	160
Data Slot length (L_{slot}) (ms)	1.024
Guard time between slots (μs)	2
Super-frame length (SFL) (ms)	20
Number of CRS in Reservation Sub-Frame	10
RTS length (bytes)	18
CTS length (bytes)	18
ResvRTS length (bytes)	23
ResvCTS length (bytes)	22
ResvConfirm length (bytes)	22
RTS mini-slot length (ms)	0.124
CTS mini-slot length (ms)	0.124
ResvRTS-mini-slot length (ms)	0.144
ResvCTS-minislot length (ms)	0.140
ResvConfirm length (ms)	0.140
CRS length (ms)	0.7
number of data slots per super-frame	12
Minimum data traffic rate (packets/super-frame)(λ_{Min})	0
Mean data traffic rate (packets/super-frame) (λ)	1
Maximum data traffic rate (packets/super-frame) (λ_{Max})	3
On period duration (t_{On}) (s)	1
Off period duration (t_{Off}) (s)	1.35
Buffer size for data traffic ($Buffer^{size}$)	10
Connection timeout (s)	10
Simulation time (s)	1000s

6.2.1.2 Impact of data traffic load on voice traffic performance

We analyze the impact of data traffic load on the voice traffic performance. The number of voice sources is set to 10, and the number of data sources is increased from 0 to 20 with an increment of 2. For the CEP scheme, we present only the impact of p^v on the dropping rate as data sources don't contend with voice sources for slot reservation during the *RSF*.

Voice packet dropping rate achieved by the RTR transmission scheme versus the increase of the number of data sources with different values of the permission probability is shown in Figure 6-2 (simulation and analytical results respectively). The voice dropping rate with the analytical model was lower than the one given by simulation, and the divergence between simulation and analytical results was 2.6 %.

Figure 6-3 shows the voice packet dropping rate achieved by the CEP scheme obtained by simulation and by the analytical model. The figure shows that simulation results agree with the results of analytical model with a divergence of 0.5 %.

In comparison between the RTR and CEP, the figures show that the dropping rate with the RTR increases linearly with the increase of the number of data sources, while it remains constant with the CEP. Except with the dynamic contention scheme where both schemes achieve low voice traffic dropping rate, we see that the voice traffic dropping rate with CEP is very low compared to the one of RTR. This is because as data sources in the CEP scheme don't compete during the *RSF* with voice sources for slot reservation, the contention rate on the *RSF* is lower than with the RTR scheme. Slot reservation for voice sources is faster with CEP scheme than with RTR scheme, and consequently, the number of voice packets dropped with CEP is lower than with RTR scheme.

With regard to permission probability, the best performance is achieved with the dynamic priority contention scheme. The low dropping rate of the dynamic priority contention scheme at high traffic load is because contending nodes with this scheme reduce their permission probability at high traffic load when they detect high collision rate during the *RSF*. Some nodes avoid contending for reservation requests transmission, the fact which contributes in decreasing the collision rate during the *RSF*. Consequently, reservations for voice sources are established faster.

Regarding the performance of the static priority scheme with different permission probabilities, the best performance are achieved by the RTR scheme with permission probabilities of (P^v=0.3, P^d=0.1). With the CEP scheme, the best performance are achieved when the permission probability of voice traffic is P^v=0.1.

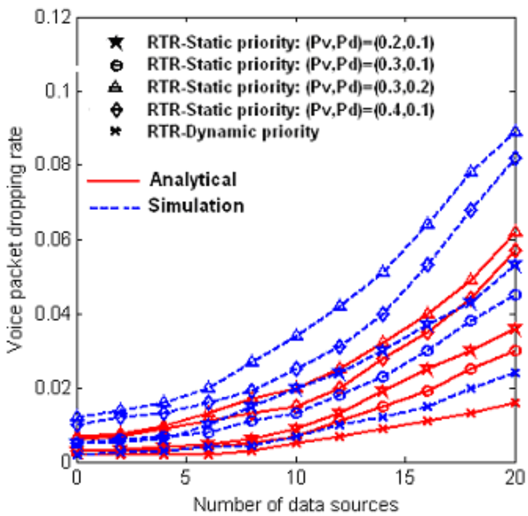

Figure 6-2. Voice traffic dropping rate vs. the increase of the number of data sources with the RTR scheme (Simulation and analytical results).

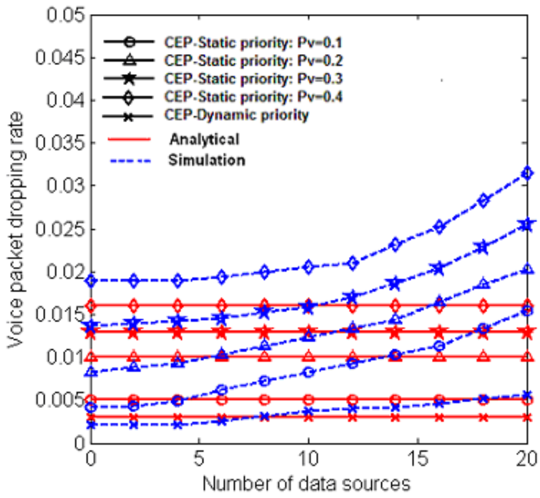

Figure 6-3. Voice packet dropping rate vs. the increase of the number of data sources with CEP scheme (Simulation and analytical results).

6.2.1.3 Impact of the voice traffic load on the data traffic performance

We analyze the impact of voice traffic load on the performances of data traffic with the RTR and CEP schemes. We consider 10 data source and increase the voice traffic load by increasing the number of voice sources. For the CEP scheme, we present only the impact of data source permission probability p^d since p^v has no effect on the data traffic performance. For the RTR scheme, we present the results in function of p^d and p^v as both parameters may have an impact on the reservation establishment delay of data sources.

Figure 6-4, Figure 6-5, Figure 6-6, and Figure 6-7 show the data traffic dropping rate and delay achieved by the RTR and CEP schemes with the dynamic and static priority contention schemes.

We observe that with the RTR scheme, p^v and p^d have not a significant effect on the data traffic delay and dropping rate. However, p^d has an important impact on the performance achieved by the CEP scheme. The CEP scheme achieves the best performance when p^d=0.1, while it achieves the highest dropping rate and delay performances when p^d=0.4.

It is observed that the RTR scheme achieves lower data traffic dropping rate and delay than CEP. The high data traffic delay and dropping rate with CEP scheme are mainly due to two reasons. The first one is packet retransmission. Indeed, as data sources compete during available slots for the transmission of each packet, packets collide frequently at high traffic load. Consequently, retransmissions are undertaken, which results in an increase of data packets delay. The second reason is the increase of packets queues length due to excessive transmission failures. Data packet transmission failures with CEP scheme contribute in queue build-up, and consequently increasing the data packet transmission delay. Packets arriving when packet queues are full are dropped.

Unlike the CEP scheme where data sources compete for the transmission of each packet, data sources with the RTR scheme transmit their packets without contention during reserved slots when slots are available and when slots are temporarily released by voice sources. Hence, data packets are transmitted faster than in the CEP and queue lengths increase slowly, which results in best data traffic performances with the RTR.

In comparison between simulation and analytical results, we can find that analytical results match simulation results for data traffic dropping rate, but not always for data traffic delay, especially at high voice traffic load. The divergence in data packet dropping rate between simulation and analytical results was 4.3 %, and the divergence in data traffic delay was 40 ms. Such a divergence can be explained by the difference between exponential traffic pattern with the simulator and its probabilistic approximation.

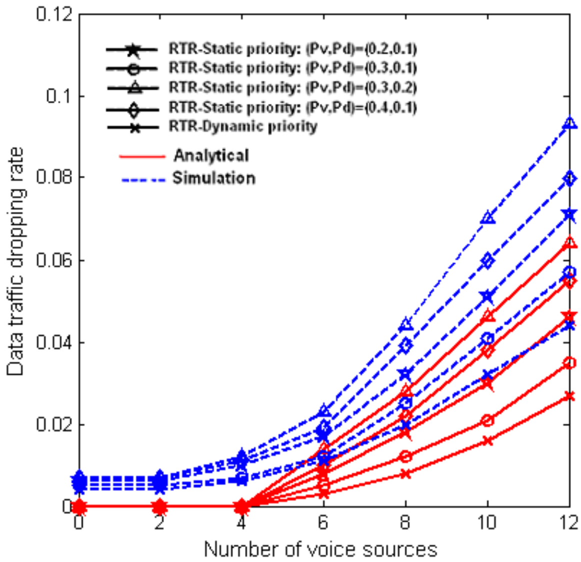

Figure 6-4. Data traffic dropping rate vs. the increase of the number of voice sources with the RTR scheme (Simulation and analytical).

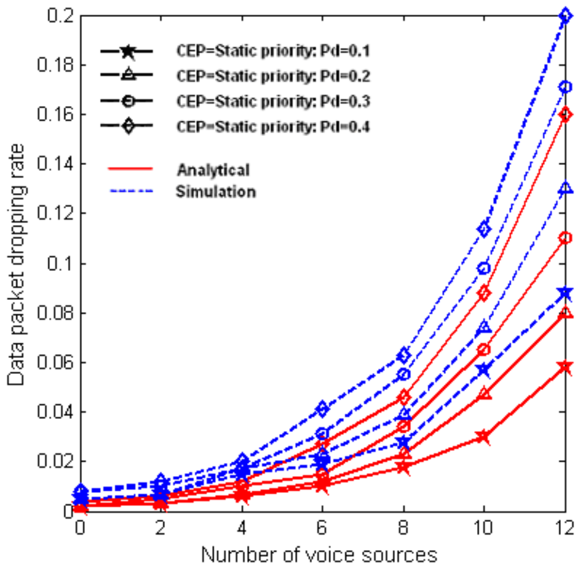

Figure 6-5. Data traffic dropping rate vs. the increase of the number of voice sources with the CEP scheme (Simulation and analytical results).

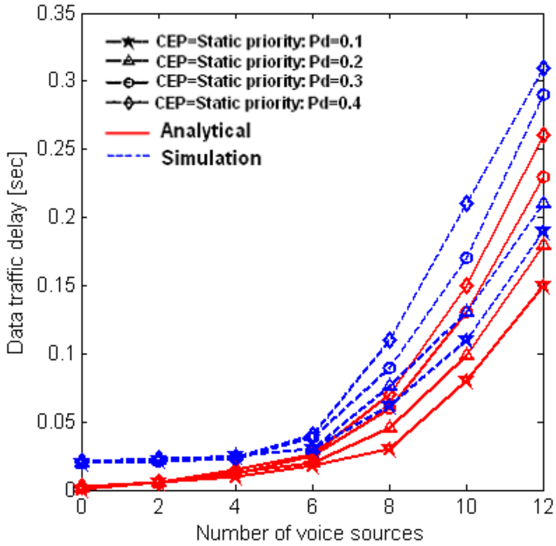

Figure 6-6. Data traffic delay vs. the increase of the number of voice sources with the CEP scheme (Analytical and simulation results).

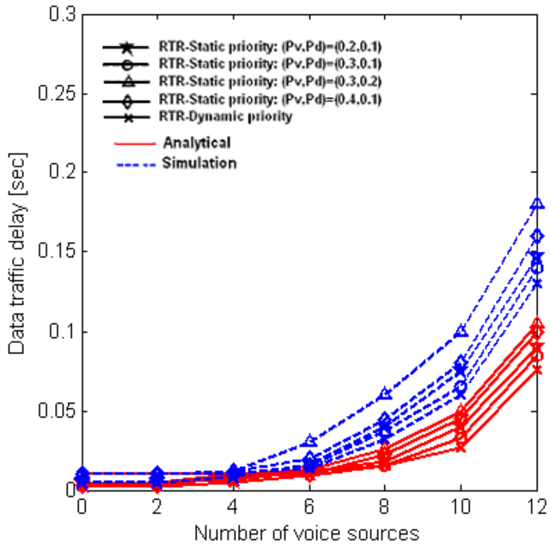

Figure 6-7. Data traffic delay vs. the increase of the number of voice sources with the RTR scheme (Analytical and simulation results).

6.2.1.4 Concluding remarks

In summary, the section highlights a good matching between analytical prediction and simulation results. The results have shown that with both CEP and RTR schemes, the dynamic priority contention scheme gives lower voice traffic dropping rate than the static priority scheme as it permits to reduce the reservation establishment delay for voice. The CEP gives better performances for voice traffic than the RTR scheme as only voice sources compete for slot reservation during the *RSF*. However, the RTR scheme gives better performance than CEP for data traffic as data sources are allowed to reserve slots and send their packets without contention.

6.2.2 Performance evaluation in a multi-hop network

In the previous section, we considered a scenario where all nodes are in the visibility of each other. The results allowed us to analyze the CEP and RTR schemes and the effect of permission probabilities and validate our stochastic model. However, this scenario does not allow the study of the performance of our protocol in the presence of hidden terminals.

In this section, we evaluate the performance of our protocol with the RTR and CEP schemes and the effect of contention schemes in a multi-hop distributed network. Our goal is to assess the effectiveness of the protocol in reducing the collision rate and providing better channel utilization at the MAC level in multi-hop distributed networks. We also compare the performance of our protocol with the IEEE 802.11e EDCF scheme. For this purpose, we use the EDCF simulation model of Wiethölter and Hoene [WH 03].

As preliminary work, we consider only unicast connections. Performance of our protocol coupled with routing for end-to-end connections will be investigated in the following chapters.

6.2.2.1 Simulation model

We consider an ad-hoc network composed of 100 nodes randomly distributed on 2000×2000 m^2 area. The transmission range of nodes is 250 meters. Each node can initiate a point-to-point G.711 voice session with one of its neighbors. Voice sessions are started at random instants of the simulation and have duration of 200 s.

For data traffic, we consider FTP sessions that transfer 10 MB files started at random instants of the simulation. Table 6-1 shows the contention parameters used for the EDCF scheme. The data traffic is associated with the background traffic class.

Table 6-1. Backoff and AIFSN values for EDCF

Traffic category	CWmin	CWmax	AIFSN
Background traffic	31	1023	7
Voice traffic	1023	15	2

6.2.2.2 Analysis of the impact of the permission probability and traffic load

We investigate the influence of the permission probabilities of voice and data sources on the performance of the CEP and RTR schemes. We consider that the 100 nodes are not mobile. We uniformly increase the traffic load by increasing data and voice sessions in equal numbers.

Figure 6-8 shows the voice traffic dropping rate achieved by the RTR scheme when (p^v,p^d)=(0.2, 0.1), (0.3, 0.1), (0.3, 0.2), and (0.4, 0.1) for the static priority contention scheme. As shown, the values taken by p^v *and* p^d have a significant impact on the performance obtained by the static priority contention scheme. As the permission probability and traffic load increase, the contention and collision rate during the *RSF* increase, and consequently, much of voice packets generated at the beginning of a talkspurt are dropped. Among the considered scenarios, the dynamic priority scheme exhibits the lowest dropping rate. The static priority scheme gives low dropping rate when (p^v, p^d)=(0.3, 0.1), while it achieves a dropping rate higher than the EDCF when (p^v, p^d)=(0.4, 0.1).

Figure 6-9 shows the voice traffic dropping rate achieved by the CEP scheme with the increase of traffic load with the dynamic priority and static priority contention schemes. The figure shows that the best performance is achieved by the CEP scheme with the dynamic contention scheme. With the static priority scheme, the best performance is achieved when p^v=0.3, while CEP behaves worst than the EDCF scheme when p^v=0.4.

We observe that the CEP scheme achieves better performance than the RTR scheme with the static priority contention scheme except when p^v=4 where both schemes achieve similar bad performance. With the dynamic priority scheme, both schemes achieve very low dropping rate in comparison with the static priority scheme.

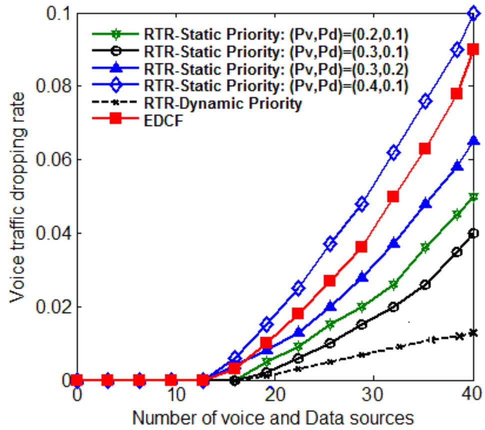

Figure 6-8. Voice traffic dropping rate vs. the increase of traffic load with the RTR scheme

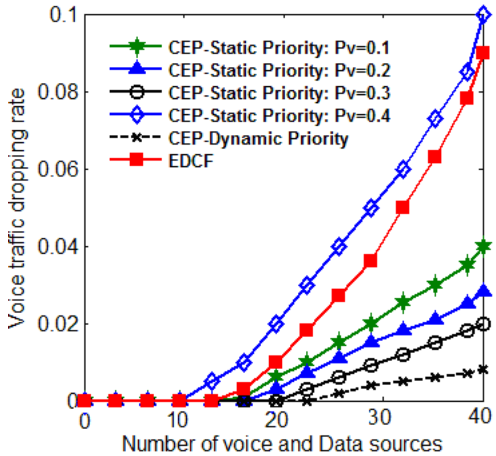

Figure 6-9. Voice traffic dropping rate vs. the increase of traffic load with the CEP scheme.

Figure 6-10 and Figure 6-11 show the data traffic throughput and delay achieved by the RTR and the EDCF scheme with the increase of voice and data traffic load. The RTR scheme exhibits the best data traffic throughput with the dynamic contention scheme, and achieves better than the EDCF. Also, the data traffic delay obtained by the RTR scheme is much lower than the one obtained by the EDCF scheme. The high delay of EDCF is because data packets are transmitted in contention with voice packets. At high traffic load, data packets experience more contention, and thus more collisions and wider backoff windows, and consequently their access delay increases. The low delay of the RTR scheme is because data packets are transmitted collision-free on reserved slots as data sources are allowed to reserve available and temporarily released slots.

Figure 6-12 and Figure 6-13 show the data traffic throughput and delay achieved by the CEP and EDCF scheme. We study only the impact of the permission probability p^d on the performance of the CEP scheme. Four scenarios are considered, namely, p^d=0.1, p^d=0.2, p^d=0.3, p^d=0.4. The *CEP* scheme achieves the best data traffic throughput when p^d=0.1, but its throughput is lower than the one of EDCF when the system becomes too congested. The data traffic throughput of CEP scheme decreases as the p^d increases, especially at high traffic load. This low throughput is due to the increase of contention rate and packet collision. At high traffic load, all slots are reserved by voice traffic and data sources send their packets mainly on temporarily released slots. With high values of p^d, data sources with queued packets contend for packets transmission more frequently, and consequently several packet collisions occur. Another factor contributing in this low throughput is the packet queue build-up. Packets arriving when packet queues are full are dropped.

It can be seen that RTR scheme achieves better throughput and lower delay than the CEP scheme.

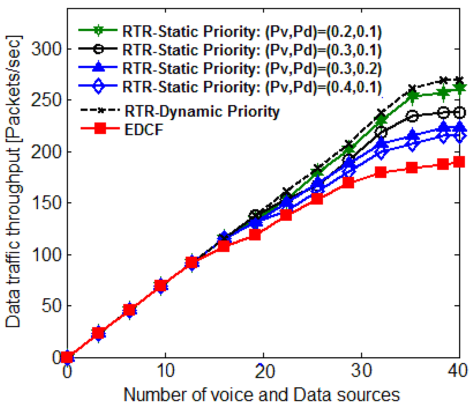

Figure 6-10. Data traffic throughput with the increase of traffic load with the RTR scheme.

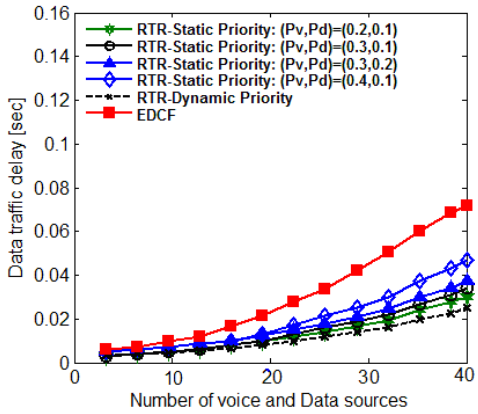

Figure 6-11. Data traffic delay with the increase of traffic load with the RTR scheme.

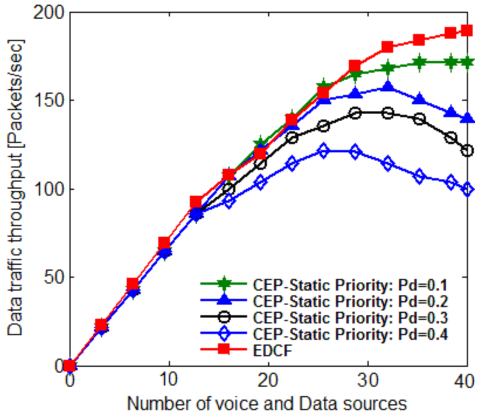

Figure 6-12. Data traffic throughput with the increase of traffic load with the CEP scheme.

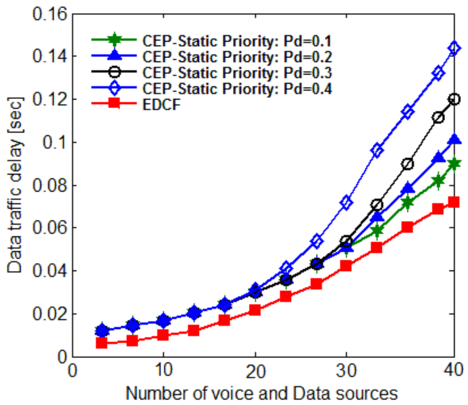

Figure 6-13. Data traffic delay with the increase of traffic load with the CEP scheme.

6.2.2.3 Concluding remarks

In this section, we evaluated the impact of the traffic load and permission probabilities on the CEP and RTR schemes in multi-hop network. As a summary, we found that with the static priority contention scheme, the CEP gives better performance than the RTR scheme for voice traffic, while the RTR scheme performs better than the CEP scheme for data traffic. However, with the dynamic priority contention scheme, the RTR scheme gives similar performance as the CEP for data traffic, and is still more efficient than the CEP for voice traffic. Thus, the RTR scheme combined with the dynamic contention scheme seems to be the best solution among the considered scenarios.

6.2.3 Impact of the slot length choice

In this section, we analyze the impact of the slot length on the performance of our reservation protocol. We evaluate the performance of the three solutions presented in section 5.2.1 for G.711, G.726, and G.729 codecs. Remember that in the first solution, the data slot length corresponds to the packet length of G.711 codec, in the second solution it corresponds to the payload of G.729 codec, while the third solution represents an enhancement of the second solution through avoiding fragmentation. We consider the RTR reservation scheme with the dynamic priority contention scheme.

Figure 6-14, Figure 6-16, and Figure 6-18 show the call acceptance ratio of voice connections with the three slot length choices. We can see that the first solution gives the same call acceptance ratio for the three considered codecs. We see also that the 2nd solution gives lower call acceptance ratio than the 1st solution for G.711 codec. This is due to the high fragmentation overhead incurred by the G.711 codec with the 2nd solution. However, the 2nd solution provides higher acceptance ratio when G.726 and G.729 codecs are used in comparison with the 1st solution. This is due to the low waste of bandwidth yielded by fragmentation with the 2nd solution with G.726 and G.729 codecs which have short packets.

The 3rd solution achieves the best call acceptance for all considered codecs because it uses short slot length, while it avoids wasting of bandwidth due to fragmentation. Indeed, G.711 codec with this solution consumes less bandwidth compared to the 2nd solution, while codecs with short packets like G.729 do not cause partial utilization of data slots.

Figure 6-15, Figure 6-17, and Figure 6-19 show the throughput achieved by the three solutions for the three codecs. The 1st solution provides the lowest throughput when G.729 codec is used. However, it achieves better throughput than the 2nd solution for G.711 traffic. Thus, the throughput of the 1st solution saturates at 16 voice connections with 500 Kbps throughput, while the throughput of the 2nd solution saturates at 14 connections with a throughput of 450 Kbps. The 3rd solution achieves the highest throughput for the three considered codecs.

The throughput achieved by the IEEE EDCF scheme for the three codecs is shown in Figure 6-20. It shows that for G.729, the EDCF outperforms our protocol when the 1st slot length choice is considered. However, our protocol outperforms the EDCF for all traffics when the 2nd or the 3rd solutions are considered.

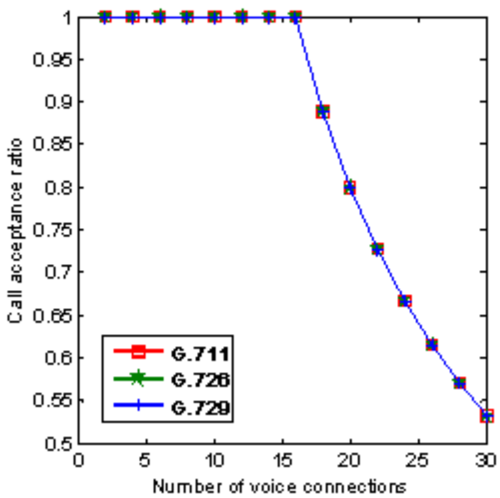

Figure 6-14. Voice call acceptance ratio with solution 1 of the slot length issue (i.e., long slot length).

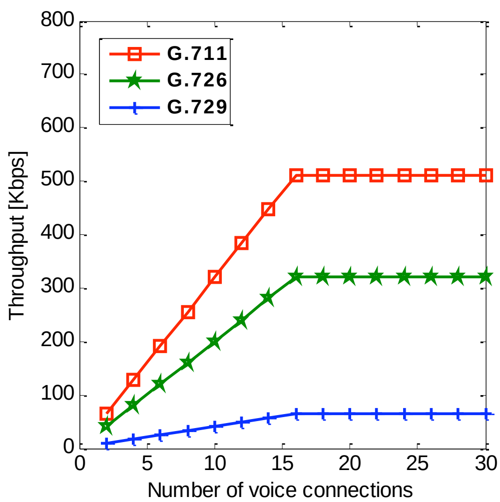

Figure 6-15. Voice traffic throughput with solution 1 of slot length issue (i.e., long slot length).

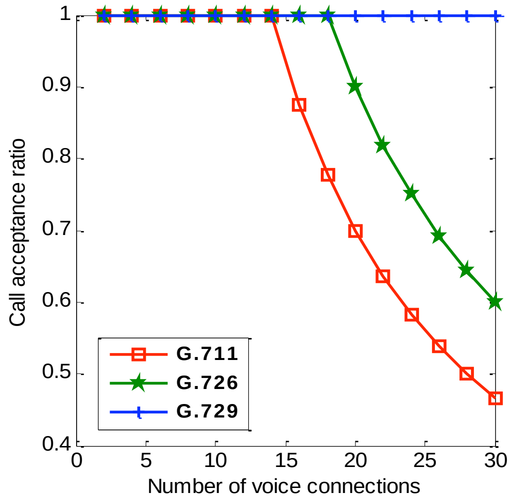

Figure 6-16. Voice call acceptance ratio with solution 2 of slot length issue (i.e., small slot length with fragmentation).

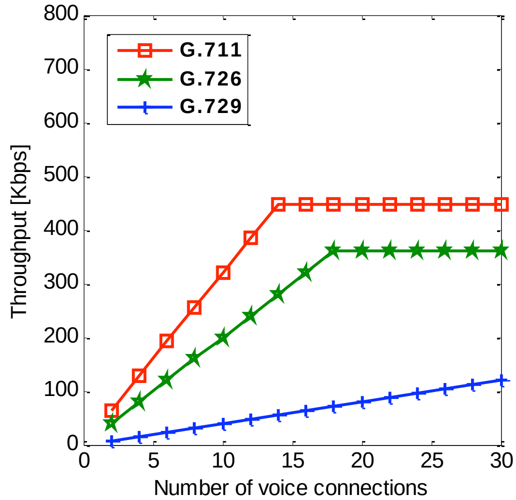

Figure 6-17. Voice traffic throughput with solution 2 of slot length issue (i.e., small slot length with fragmentation).

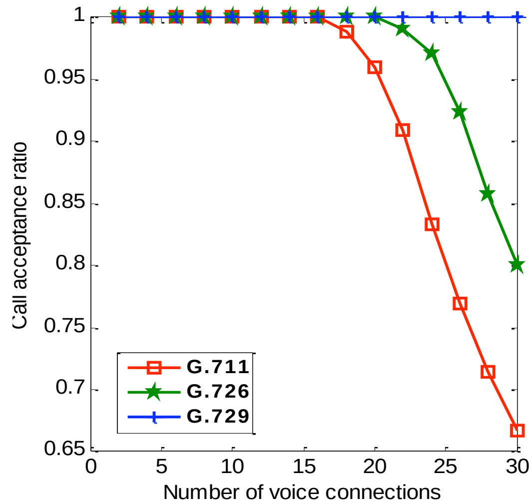

Figure 6-18. Voice call acceptance ratio with solution 3 of slot length issue (i.e., small slot length without fragmentation).

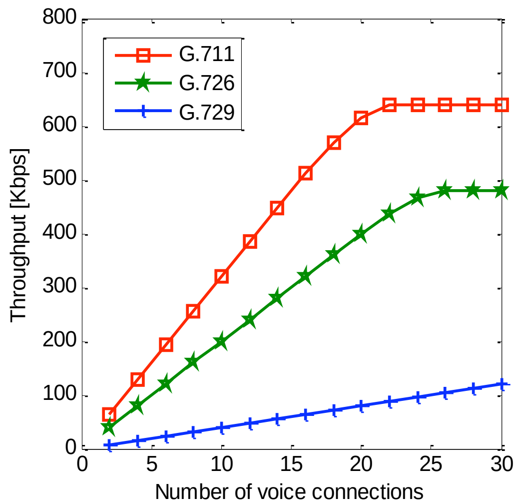

Figure 6-19. Voice traffic throughput with solution 3 of slot length issue (i.e., small slot length without fragmentation).

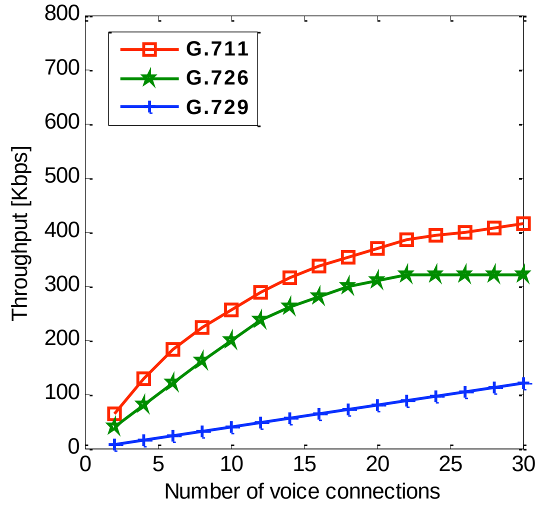

Figure 6-20. Voice traffic throughput with EDCF.

6.2.3.1 Concluding remarks

We evaluated the impact of the slot length on the performance of our reservation protocol. As a summary, we found that a choice of short slot length is suitable for codecs with small packets, but incurs a significant waste of bandwidth for codecs with long packets due to fragmentation overhead. A choice of long slot length is more suitable for codecs with long packets, but may lead to under-utilization of data slots when codecs with short packets are used. An alternative solution consists in choosing small slots while avoiding fragmentation overhead through allowing voice sources with long packets to reserve adjacent small slots, and sending voice packets without fragmentation. Simulation results show that this solution is more efficient than the two other solutions. Our solution is more efficient than the EDCF scheme.

6.3 Conclusion

In this chapter, we analyzed in detail the performance of our solution for bandwidth reservation for voice traffic support over MANETs (cf. Chapter 5). The analysis is performed through a stochastic model as well as simulations. The stochastic model along with simulations allowed us to dimension efficiently our solution. Thus, it allowed us to analyze the impact of the protocol parameters such as the permission probabilities, to compare static and dynamic priority contention schemes, and to compare the behavior of our protocol with the RTR and CEP schemes (cf. Chapter 5).

Both analytical and simulation results show that the dynamic priority contention scheme performs better than the static priority contention scheme, and is less sensitive to the traffic load. Results show also that the RTR scheme is more efficient than the CEP scheme for both voice and data traffic, especially when combined with the dynamic contention scheme.

When the slot length is of concern, the results show that the best performances are achieved when considering small slots, and allowing traffic sources with long packets to reserve several contiguous slots without fragmentation.

In the next step of our work, we focus on another issue with reservation MAC protocols in MANETs, which is mobility of nodes. Indeed, we extend our protocol proposed in chapter 5 in order to establish end-to-end reservation along a path.

Chapter 7: DSR-based End-to-End Bandwidth Allocation Scheme for Voice Traffic Support

7.1 Introduction

Until now, we focused on problems related to bandwidth reservation on point-to-point links. In fact, despite an efficient point-to-point reservation MAC scheme is a primary requirement for QoS provisioning, reserving resources along a path is also of great importance. Most reservation protocols proposed in the literature focus on point-to-point reservations, and only few work has been done to propose an efficient end-to-end reservation scheme. In fact, the task of end-to-end bandwidth reservation cannot be efficiently fulfilled without providing a tight coordination between the MAC sub-layer and the routing protocol.

Another issue with reservation protocols that we consider in this chapter is mobility of nodes. When mobility of nodes is of concern, new challenges appear when designing reservation protocols. In some scenarios, mobility of nodes causes some reservation being broken without a significant change in the end-to-end paths. Such scenarios can be dealt with by the MAC sub-layer without any operation from the routing level. In other scenarios, mobility of nodes may cause reservation breakage as well as significant changes in network topology. Both scenarios result in significant performance degradation, and should be considered at MAC and routing levels.

In this chapter, we propose a reservation scheme called End-to-End Reservation scheme for Voice and data traffic support (EERV) which is an extension of ARPV (cf. Chapter 5) to support the reservation and release of resources along a path in cooperation with the routing layer. In addition to end-to-end bandwidth reservation, EERV includes mechanisms that alleviate performance degradation due to mobility of nodes. These mechanisms are mainly reservation loss detection and reservation recovery.

7.2 State of the art

Despite the numerous reservation protocols proposed in the literature, only few work considered the problem of end-to-end resource reservation.

MACA with Piggy-backed Reservation (MACA/PR) [LG 97] is a protocol used to provide real-time traffic in multi-hop wireless networks. The main components of MACA/PR are: a MAC protocol, a reservation protocol, and a QoS routing protocol. Time is divided into slots of varying lengths, which are asynchronous between nodes. Each node in the network maintains a reservation table that records all the reserved transmit and receive slots/windows of all nodes within its transmission range. QoS routing protocol used with MACA/PR is an extension of Destination Sequenced Distance Vector (DSDV) routing protocol [PW 94], where bandwidth constraint has been introduced in the routing process. Each node broadcasts to its neighbors the (bandwidth, hop distance pairs) for each destination and for each bandwidth value. Real-Time Medium Access Control protocol (RTMAC) [MS 02, BMS

04, and MVS 04] is very similar to MACA/PR as it does not need global slot synchronization, and as it extends DSDV for routing. With both of MACA/PR and RTMAC, QoS routing requires the piggy-backing of routing packets with reservation tables.

The major advantage of MACA/PR and RTMAC is they do not need synchronization. However, they suffer the important control overhead inherited from DSDV. Thus, the DSDV protocol requires the periodic exchange of routing messages in order to calculate the shortest paths, leading to excessive bandwidth consumption. Furthermore, in addition to the simple information about routing, routing messages with MACA/PR and RTMAC include reservation tables leading to excessively long messages, which increases further bandwidth consumption.

In DARE [CAR 05, CAR 06], authors propose a protocol for bandwidth reservation along a path in multi-hop networks. Authors assume that there is a routing protocol, precisely AODV, providing the reservation protocol with routes between the source and the destination. To establish an end-to-end reservation, the source sends a Request-To-Reserve (RTR) message including the periodicity and duration of the reservation. Each intermediate node along the path which receives the RTR forwards the RTR to the next-hop if it can accept the reservation request. When the destination receives the RTR it generates a Clear-To-Reserve (CTR) along the reverse path to indicate to all intermediate nodes that the reservation request is accepted.

Compared to MACA/PR and RTMAC, DARE has the advantage of low control overhead as it does not need periodic exchange of routing messages in order to find paths. However, despite its low overhead DARE suffers two main problems. The first one is that it does not ensure consistency of reservations (cf. Chapter 3) as there is no guarantee that reservation control packets (RTR and CTR) are received by all one-hop neighbors of intermediate nodes. The second issue is that it considers only the shortest path provided by the routing protocol. Thus, there is no guarantee that this path provides the required bandwidth. Thus, depending on the actual reservations in the network, some nodes may not be able to fulfill the required reservation.

7.3 Problem statement

The design of solutions for end-to-end bandwidth reservation needs to consider at least the three following functions:

- *Efficient reservation MAC protocol:* the MAC sub-layer enables nodes to establish point-to-point reservation and resolve contention between nodes during bandwidth reservation phase. Thus, all issues related to point-to-point reservation such as reservation conflicts, collision of packets, and medium access need to be handled by the MAC sub-layer. This was the focus of previous chapters.

- *Efficient end-to-end bandwidth reservation:* The second important feature in the design of the end-to-end reservation scheme is how to extend the reservation MAC protocol in order to provide efficient end-to-end reservation. Thus, this objective can be achieved only if a judicious choice of the routing protocol and its coordination with the MAC sub-layer are provided. Actual solutions for end-to-end bandwidth reservation base their reservations on the path provided by the routing protocol which is generally the shortest path, and assume a reservation failure if this path does not provide the required bandwidth. However, when designing an end-to-end bandwidth reservation solution, bandwidth availability becomes an important path selection criterion in addition to the path length. Thus, there may exist paths other than the shortest one that provide the required bandwidth.

- *Efficient mobility handling:* One critical phenomenon of paramount importance in reservation protocols is the reservation clash [BTM 08-1, BTM 08-2]. The reservation clash happens when two nodes which are far away from each other and which have reserved the same slot move. If one of them enters in the transmission range of the other, collisions happen in reserved slots and one (or both) of them loses its reservation. Reservation clash have drastic consequences on the QoS provided to real-time applications. Reserving nodes affected by reservation clash suffer excessive packets collisions and dropping. Thus, a reservation protocol should provide mechanisms for reservation clash detection and its handling.

The end-to-end reservation scheme presented in the rest of this chapter has as primary aim to ensure these three requirements.

7.4 End-to-End Reservation scheme for Voice and data traffic support (EERV)

7.4.1 Assumptions

We consider the following assumptions:

- At the MAC sub-layer, we consider the use of ARPV protocol with the RTR (Reserve Temporarily Released slots) scheme and the dynamic contention scheme (cf. Chapter 5).

- Links are symmetric.

7.4.2 End-to-end bandwidth reservation

In addition to efficient reservation MAC protocol, end-to-end resource reservation requires the design of modules at the network layer that handle routing and end-to-end resource reservation. Indeed, two primary modules are required at network layer: routing and reservation modules.

The reservation module at the network layer should do abstraction of all problems related to medium access (such as contention resolution, packet collision, slot reservation, and packet retransmission) which are considered resolved by the MAC sub-layer.

Figure 7-1 shows the main modules involved in end-to-end bandwidth reservation, and the interactions between these modules. At the MAC sub-layer, we consider the use of our reservation protocol described in chapter 5 with the RTR scheme and dynamic contention schemes. Next, we give the main operations implemented by the routing protocol and reservation module, and the interaction between them and with the MAC sub-layer.

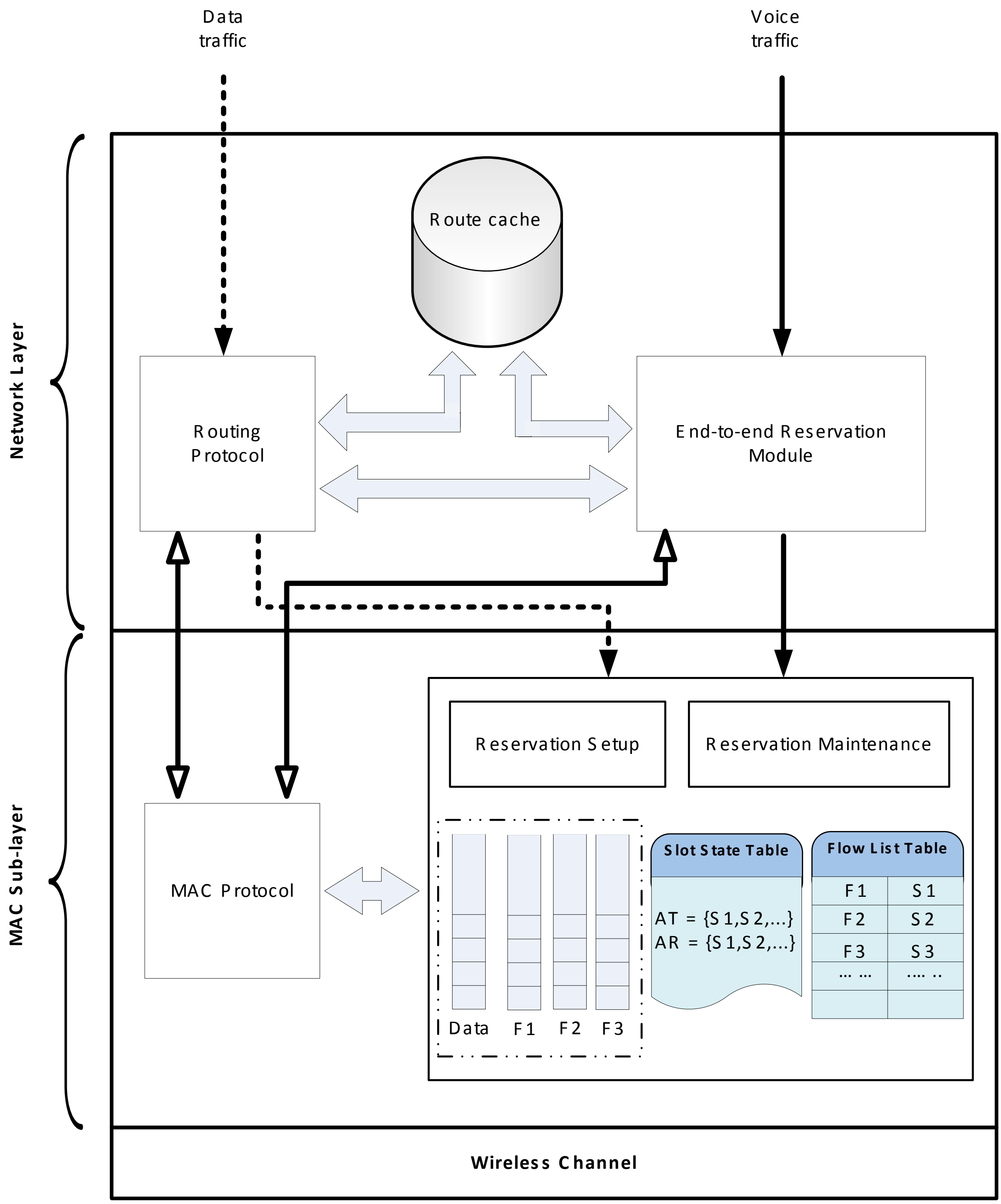

Figure 7-1. Routing and reservation modules for end-to-end resource reservation.

7.4.2.1 Routing module

The routing module determines routes between a source and a destination, and forwards packets to the next intermediate relay nodes. In order to operate properly a routing protocol should implement the three following tasks [ESP 08]:

- *Routing information dissemination:* It consists in the exchange of routing information between nodes forming the network, and provides information on the network topology and the paths that drive toward the destination. Depending on the quantity of information exchanged, nodes can obtain a more or less accurate view of the network topology. Thus, the routing protocol should reduce the amount of routing information exchanged in order to optimize bandwidth utilization.

- *Route selection:* Based on the information collected of the network topology, the routing protocol chooses a path among the feasible paths based on some criterion. A common criterion in path selection is minimizing the number of hops. However, other important criterion such as maximizing bandwidth on the selected path to satisfy QoS or maximizing the residual bandwidth should be considered when bandwidth reservation is of concern.

- *Route maintenance:* the routing protocol should react quickly to topology changes. New paths should be found rapidly so that path breakage does not cause drastic performance degradation. Furthermore, finding new paths should not incur a huge control overhead.

In fact, there is tradeoff between fast route discovery and control overhead. On one hand, proactive routing protocols (such as DSDV [PW 94] and OLSR [RFC 3626]) assume periodic exchange of routing messages. Each node maintains a routing table that contains information (for example the next-hop node on the path) necessary to reach any other node in the network. Hence, proactive protocols react faster to topology changes. However, depending on the frequency of exchange of routing messages, proactive protocols can result in a huge control overhead. In addition, routing messages are exchanged even when no new route discovery is required. On the other hand, reactive (or on demand) protocols (such as AODV [RFC 3561], DSR [RFC 4728, JMB 96, JMB 01]) exchange routing information only when a new route toward the intended destination is required. Compared to proactive protocols that maintain routes toward all nodes of the network in their reservation tables, reactive protocols maintain only useful routes. As a result, the size of routing tables is low compared to proactive protocols.

Finding a route on demand in reactive protocols creates some delay before the route is found. This delay is incurred only at the time of route discovery or when a path is broken. While waiting for route discovery, packets received from upper layer are queued.

In our reservation scheme, we consider a reactive routing protocol, namely the Dynamic Source Routing (DSR) [RFC 4728, JMB 96, JMB 01] protocol. This choice is motivated by the low control traffic overhead of this protocol. Another motivation of this choice is that DSR maintains in its route cache several paths to any destination, and allows each sender to select and control the routes used in routing its packets, for example, for use in load balancing or for increased robustness [RFC 4728]. Thus, it is possible to explore alternative paths to achieve certain objective. In our end-to-end reservation scheme, we use information in the route cache in order to explore and reserve resources along the path that provides the required bandwidth (cf. Section 7.4.6).

7.4.2.2 Reservation module

The reservation module is used to establish, adapt, restore, and tear down reservations along a path once a path is found by the routing protocol. In our scheme, when the reservation module wants to reserve resources along a path for a traffic flow, it transmits a *RESV_REQ* packet. This packet is used as a means to request intermediate nodes along the path to reserve resources. However, we make a distinction between the reservation for voice traffic and data traffic.

7.4.2.2.1 End-to-end reservation for voice traffic

Before starting the reservation process, the routing protocol is involved in finding a path. The route discovery mechanism of DSR allows nodes to find a route between the source and the destination. Notice that for voice connections, voice packets are generated by the source only once the end-to-end reservation is established.

Once the route is found, the reservation module of the source node requests the MAC sub-layer to establish a point-to-point reservation with the next-hop node. If the point-to-point reservation is successfully established, the reservation module sends a *RESV_REQ* message to the destination. The *RESV_REQ* specifies the flow identifier and the list of nodes on the path to which the *RESV_REQ* should be forwarded. This path is obtained from the routing module. Then, the *RESV_REQ* is sent on the slot which is reserved by the MAC sub-layer to the traffic flow. Each node along the path that receives the *RESV_REQ* delivers it to the reservation module of the node. The latter checks the identifier of the next node on the path, and requests its MAC sub-layer to establish a point-to-point reservation with this node. If the reservation is successfully established, the reservation module forwards the *RESV_REQ* to the next-hop node. Otherwise, a *RESV_ERROR* is sent to the reservation module of the source node. Slots reserved on the sub-path from the source to the node which could not establish a reservation are released through the reservation release scheme described in section 7.4.3. Afterwards, the reservation module of the source repeats the end-to-end reservation process up to a *maximum_reservation_retry_limit* times after which the traffic flow is rejected, and the application layer is informed of the connection failure.

If reservations are successfully established on all nodes along the path, the *RESV_REQ* reaches the destination. Thus, each MAC entity along the path has registered the slot which is reserved to the traffic flow. The reservation module of the destination sends a *RESV_REP* message to the source to confirm the end-to-end reservation. The *RESV_REP* message travels on the reverse path toward the source. When the reservation module of the source node receives the *RESV_REP*, it informs the application layer about the connection, and starts sending voice packets to the MAC sub-layer which sends them on the reserved slot. Finally, each node along the path which receives a voice packet checks the flow identifier of the packet, and sends it on the slot which is reserved to that traffic flow. This process is repeated until the packet reaches the destination.

An example of end-to-end reservation scenario is given in Figure 7-2 where node B wants to establish an end-to-end voice connection with node E. For sake of simplicity, we put only the *ResvRTS/ResvCTS/ResvConfirm* handshake for point-to-point reservation.

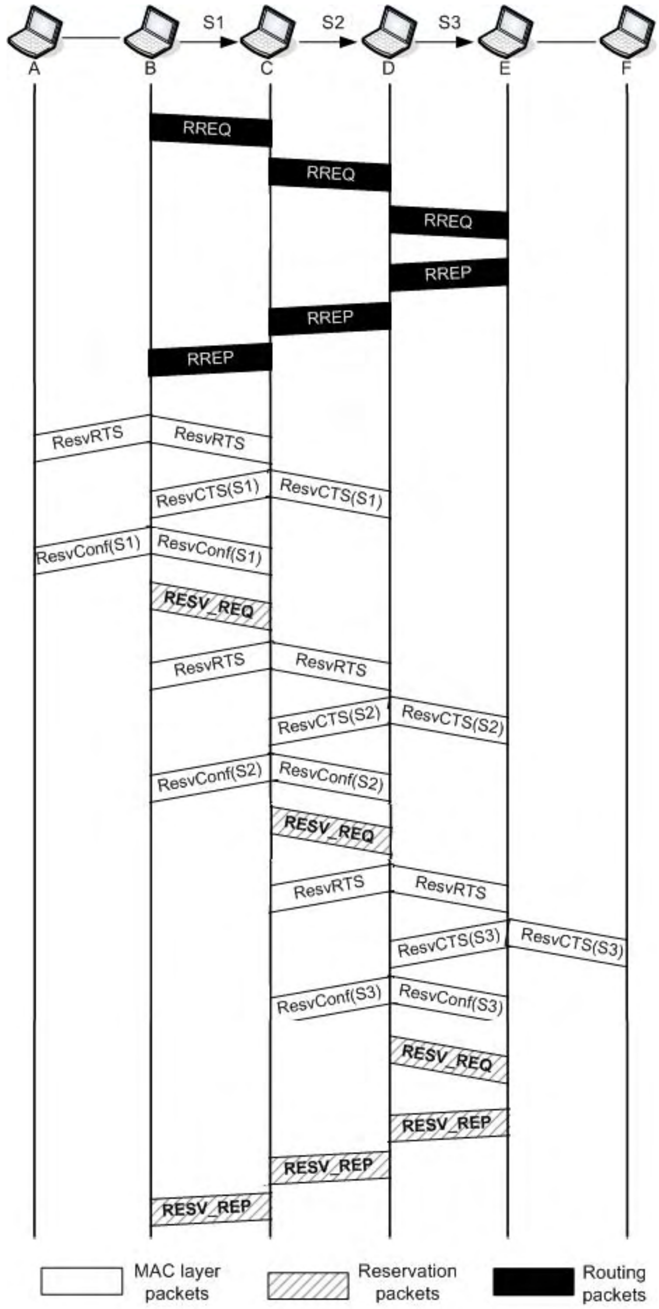

Figure 7-2. End-to-end reservation establishment scenario.

First, B initiates the route discovery to find a route towards E. Once the route found, the reservation module of node B requests the MAC level to reserve a slot with node C. B establishes a reservation with C on slot S1 by sending a *ResvRTS* to node C. Node C responds with a *ResvCTS* accepting the reservation on slot S1, and node D records the reservation. Node B confirms the reservation, and node A records the reservation of slot S1 when it receives the *ResvConfirm*. Once the reservation established between B and C, a *RESV_REQ* is sent by B to C. When the *RESV_REQ* arrives to C, the MAC sub-layer is requested to establish a reservation with D. Slot S2 is reserved between nodes C and D, and nodes B and E record the reservation. After, node C forwards the *RESV_REQ* to node D. Node D establishes a reservation on slot S3 with E, and nodes C and F record the reservation preventing collision during the reserved slot. Then, node D forwards the *RESV_REQ* to node E which sends a *RESV_REP* to B on the reverse path E-D-C-B. Once node B receives the *RESV_REP*, packets start travelling from the source node B to the destination D on slots S1, S2, and S3 collision-free.

7.4.2.2.2 Bandwidth reservation for data traffic

We consider that data sources have no specific delay requirements, and data packets can be queued at intermediate nodes. Thus, for data traffic, only the routing module is involved in end-to-end transmission. The reservation module is not used because unlike voice traffic where bandwidth should be available along the path, we consider that for data traffic, packets are queued if some nodes along the path have not enough bandwidth. Thus, any path found by the routing module can be used for data packet delivery regardless of slot availability on this path.

When data packets are generated at the source node, first, the routing protocol searches for a path towards the destination if no route is available in the route cache. After, the routing protocol starts forwarding data packets on the path. At the MAC level, each node along the path that receives a data packet tries to reserve a slot with the next-hop if a reservation for the data traffic is not yet established. If some node of the path cannot establish a point-to-point reservation, it queues the packet and tries to establish a point-to-point reservation later. If the node retries the reservation for the *maximum_data_pkt_retransmission* times and no reservation confirmation is received, the node drops the packet. We assume that data traffic is issued by TCP connections and hence packet loss detection and retransmission are provided by TCP [TCP].

7.4.3 End-to-end reservation release

End-to-end reservation release is an important component of the reservation protocol since it allows restoring slots which are no longer used for transmission along the path. It is needed in the two following cases:

7.4.3.1 Silent voice source has temporarily released its reserved slot

As illustrated in chapter 5, a voice source is considered in a silent period when its packets queue is empty. We showed that at the MAC level, if a node which reserved a slot for voice does not transmit a packet in its reserved slot, the slot is considered temporarily released. Neighbor nodes mark the slot as temporarily transmission released, and are allowed to reserve temporarily this slot for data reception. The same way, if a node which receives voice packets on a slot does not send an ACK on the ACK mini-slot of its reserved slot, the slot is considered temporarily reception released, and neighbor nodes are allowed to reserve this slot for transmission.

Temporarily reservation releases should be propagated to all nodes of the path. Thus, when the voice source (first node of the path) switches to the silence period, it temporarily releases its reserved slot with the second node on the path. The reservation release is propagated to all nodes of the path when each one of them does not receive (or does not transmit) voice packets on its reserved slot. Thus, each node adjacent to the path updates its Slot State Table, and mark the slot temporarily released.

When a voice source switches to the activity period, it restores its reservation with the second node on the path. The voice source starts sending its packets on its reserved slot when the reservation is restored. When each node along the path starts receiving voice packets again on its reception reserved slot, it restores the reservation which has been temporarily released for the traffic flow. Thus, reservations are restored on all nodes on the path for the traffic flow and adjacent nodes are aware that the reserved slots are no more temporarily released.

7.4.3.2 End-to-end reservation failure

As illustrated in section 7.4.2.2.1, the end-to-end reservation process for voice traffic fails if one of the nodes on the path cannot establish a reservation. Thus, reservations that were reserved on this path should be released as they do not form a valid reservation. These reservations are released in the same way as in the reservation release scheme for silent voice sources. As the end-to-end reservation is not completed, voice packets will not travel along the path. Consequently, point-to-point reservations which were established along the path are released when the *end_of_connection* timeout associated to these reservations expires.

7.4.4 Reservation breakage due to mobility of nodes

Mobility of nodes may cause either local reservation loss or path breakage. The first issue occurs when conflicts of reservations appear at some nodes on the path without any change in the end-to-end path. The second issue occurs when mobility of nodes causes significant changes of topology and routes, and searching new routes is required.

7.4.4.1 Local reservation loss without path breakage

Local reservation loss is detected by a point-to-point reservation receiver, and is handled by the MAC sublayer. A reservation on slot t is lost at a receiver R if several neighbors of node R reserve slot t for transmission. This happens when R moves toward another node which has reserved the same slot t for transmission, or when some node which reserved slot t for transmission moves toward the receiver R.

If node R detects collision during the reserved slot for *loss_detection* times successively, it concludes that the reservation is lost. The receiver applies the *local reservation recovery* process which consists of two steps. The first step consists in releasing the lost slot with the node preceding node R on the path through sending a *ResvRelease* packet. The second step consists in the *local reservation repair*. When the node preceding node R on the path (denote it E) receives the *ResvRelease*, it cancels its reservation and stops sending packets during the reserved slot. Packets waiting for reservation repair are queued until a new reservation is established. Afterwards, node E restarts the reservation negotiation process with node R in order to reserve another slot.

Through this scheme, the MAC sub-layer handles local reservation loss and mobility of nodes without any action from upper layer modules (if the mobility does not cause path changes).

To illustrate the local reservation recovery mechanism, we consider the scenario of Figure 7-3 where two end-to-end reservations are established between nodes S1 and D and between S2 and D. If node A moves towards node F and stays in the neighborhood of S2 and E, packets transmitted by node C will collide with packets of node A on slot 2, while packets transmitted by node A will be correctly received by E. Only nodes F and C suffer reservation breakage. The same path and reservations can be kept, but the reservation of node C on slot 2 should be changed. Node C releases its reservation with node F on slot 2 and establishes a new reservation on another slot.

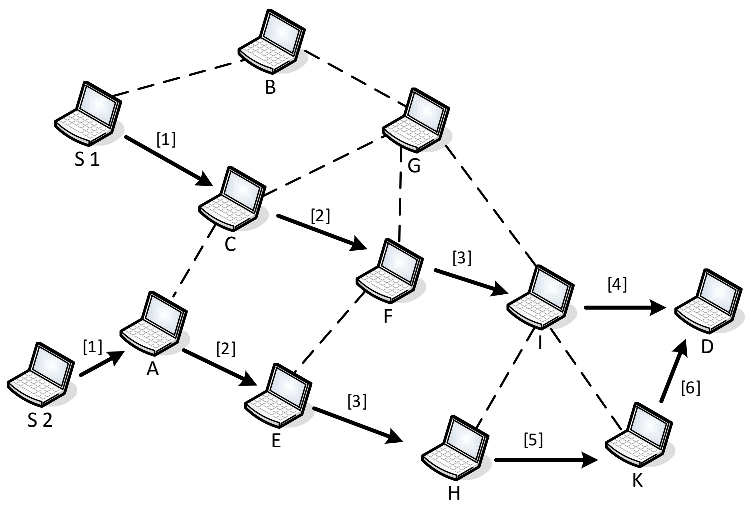

Figure 7-3. Local reservation loss and local reservation recovery (dashed lines represent logical links. The number above a link denotes the slot reserved on the link).

7.4.4.2 Reservation loss due to path breakage

Unlike the scenario described in the previous section, mobility of nodes (or node failure) may result in path breakage. In this case, reservations should be released on all nodes forming the old path, and a new path should be found and new reservations along this path should be established.

Path breakage at the routing level is detected by the route maintenance mechanism when a node on the path is unable to forward a packet to the next-hop node. Thus, each packet is retransmitted (up to a *maximal_number_of_attempts*) until a confirmation of receipt is received. If a packet is retransmitted by some node on the path *maximal_number_of_attempts* times and no receipt confirmation is received, the routing module of the node returns a *ROUTE ERROR* message to the source node.

For example, consider the scenario of Figure 7-4 where an end-to-end reservation is established between nodes S and D along the path S-C-F-I-D. Suppose that node F is switched off. If C is unable to deliver a packet to the next-hop F, then C returns a *ROUTE ERROR* to S. If S has another route to reach D in its route cache (for example suppose node S has the route S-A-E-H-K-D in its cache), this route is considered. Otherwise, S initiates a new Route Discovery phase to find another route to reach D.

In conjunction with this route maintenance scheme, a reservation recovery scheme should be applied by EERV to release slots reserved on the old path, and establish a new end-to-end reservation on the new path.

When an up-to-date route is found, node S initiates a new end-to-end reservation phase following the steps described in section 7.4.2, and a new reservation along the path S-A-E-H-K-D is established.

Nodes on the sub-path from the node following the failed node to the destination are only required to release the slots which were reserved for the traffic flow for transmission and reception on the old path. These slots are released when clear channel is detected during the data slot and the ACK mini-slot of these slots during *end_of_connection* period of time. Consider the scenario of Figure 7-4 where node F is switched off. Node I and neighbors of node F (i.e. C, G, and E) released slot 3 when they detect clear channel during this slot. As node I will not transmit acknowledgment on the ACK mini-slot of slot 3, neighbors of node I (i.e. H, K, and D) will release this slot. Similarly, when node I stops forwarding packets on its reserved slot 4, this slot will be made available for reception at H and K. D stops sending acknowledgement on slot 4 and the slot is made available for transmission at K.

The other nodes (nodes on the sub-path from the source to the node preceding the broken link) will release their reservations when the source node stops sending packets on the old path. May be some of these nodes will be part of the new path. These nodes will maintain slots which have been already reserved to the traffic flow if these slots have not been yet released.

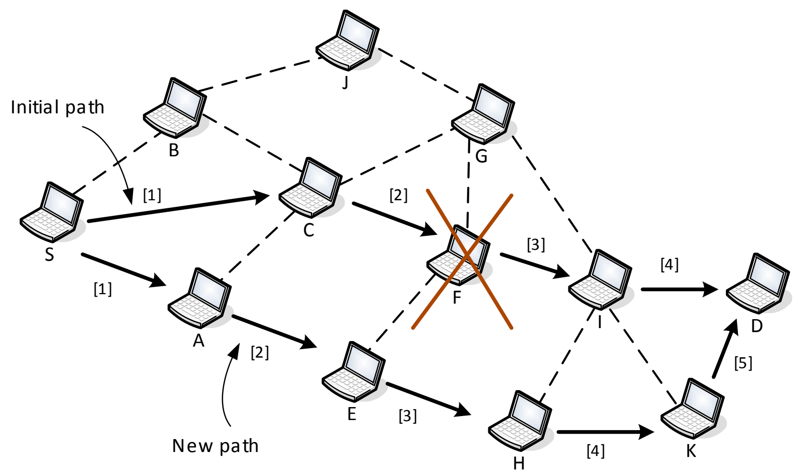

Figure 7-4. Path breakage and reservation recovery scenario.

However, packets transmitted on the sub-path from the source to the node preceding the broken link before a new path is found will be queued and dropped when they reach the broken link. Since the voice source is not considered to reissue its voice packets this is a serious problem. This issue can be avoided through the salvaging scheme of DSR [RFC 4728]. Consider the path breakage scenario in Figure 7-4. After sending the *ROUTE ERROR* to node S, node C searches for a route from itself to the destination D. This route may be cached or found by a new route discovery. If such a route is found (route C-G-I-D in Figure 7-4), node C replaces the original route on the queued packets with its own route and forwards the queued packets on this route. After, the reservation module of node C establishes reservations on this route (i.e. C-G-I-D), and queued packets are sent on this route. If this route and the old path have some nodes in common (such as nodes I and D in our example), these nodes will keep the slots that they have already reserved for the traffic flow if these reservations have not yet expired. However, these reservations are only to serve packets queued at node C and are released when the source node S finds a new path.

7.4.5 Packet re-routing and reservation swinging

In some scenarios, mobility of nodes leads to topology changes without causing reservation or path breakage on the current path. Thus, the source node routing layer may discover new shortest paths during the connection. As described in RFC 4728, a node adds new paths to its Route Cache as it learns of new links between nodes in the ad hoc network.

An alternative shortest path to reach a destination D1 can be discovered in three cases. The first one is when the source node applies the route discovery to reach another destination D2 where D1 belongs to the path towards D2. The second case is when the source node forwards a packet (data or *ROUTE REQUEST* or *ROUTE REPLY* packet) where the destination D1 is among the nodes to which the packet is forwarded. The third case is when nodes operate their interfaces in *promiscuous* mode, disabling the interface address filtering and causing the network protocol to receive all packets that the interface overhears. These packets allow nodes to learn potentially useful information for routing, while causing no additional overhead on the limited network bandwidth.

In the event the source node routing entity discovers a new shortest path, the reservation module tries to reserve resources along the new path. Reservations are established on the new path in the same way as described in section 7.4.2.2.1. If reservations are successfully established on all nodes of the new path, the source stops sending packets on the old path, and slots reserved on the old path are released when clear channel is detected during the data slot and the ACK mini-slot of these slots (see section 7.4.3). Otherwise, reservations on the old path are kept and packets continue being transmitted over this path. Thus, new discovered shortest paths (if any) are taken into account in EERV only if resources are available on these paths.

This reservation swinging scheme results in a decrease of the bandwidth consumed by traffic flows as the number of hops and slots allocated on the new path are less than the ones on the old path. Moreover, the end-to-end delay is reduced, which is useful for delay sensitive traffic.

7.4.6 End-to-end reservation in the presence of multiple paths

The Route Cache in DSR supports storing more than one route to each destination (see RFC 4728). In the case where several paths are available in the Route cache (at path breakage or new path discovery), the reservation module of EERV tries to reserve resources along the shortest path first. If resources are available on this path, this path is considered for the traffic flow, resources are reserved and packets are forwarded on this path. Otherwise, the other paths available in the Route Cache towards the same destination are explored from the shortest to the longest.

To illustrate this scheme consider the scenario of Figure 7-5, Figure 7-6, Figure 7-7, and Figure 7-8. In this scenario node 1 have a traffic flow to send to node 5, and wants to establish an end-to-end reservation with node 5. We assume that the traffic flow requires one slot per super-frame. First, node 1 applies the route discovery by sending a *RREQ* destined to node 5 (cf. Figure 7-5). As shown in Figure 7-6, node 1 receives a *RREP* from the destination, node 5, and from node 3 which has a route to reach node 5 in its Route Cache. Consequently, node 1 has two paths to reach node 5 in its Route cache. Then, node 1 tries to reserve resources along the shortest path, i.e., 1-2-4-5 (cf. Figure 7-7). The *RESV_REQ* travels up to node 4 which tries to reserve a slot with node 5 at the MAC level. As there are no common available slots between nodes 4 and 5 the point-to-point reservation fails, and node 5 do not receive the *RESV_REQ*. Consequently, no *RESV_REPLY* is issued by the destination, node 5, and the end-to-end reservation process on this path is considered failed. Afterward, node 1 tries to establish an end-to-end reservation on the second path toward node 5 in its Route Cache, i.e., 1-2-3-6-5 (cf. Figure 7-8). The *RESV_REQ* travels along the path 1-2-3-6-5 and each node on the path reserves a slot and forwards the *RESV_REQ* to the next-hop node. As all nodes along the path are able to reserve a slot for the traffic flow, the *RESV_REQ* arrives to node 5 which sends a *RESV_REPLY* back to node 1 on the reverse path 5-6-3-2-1. Thus, the end-to-end reservation is successful and node 1 starts sending its packets on the path.

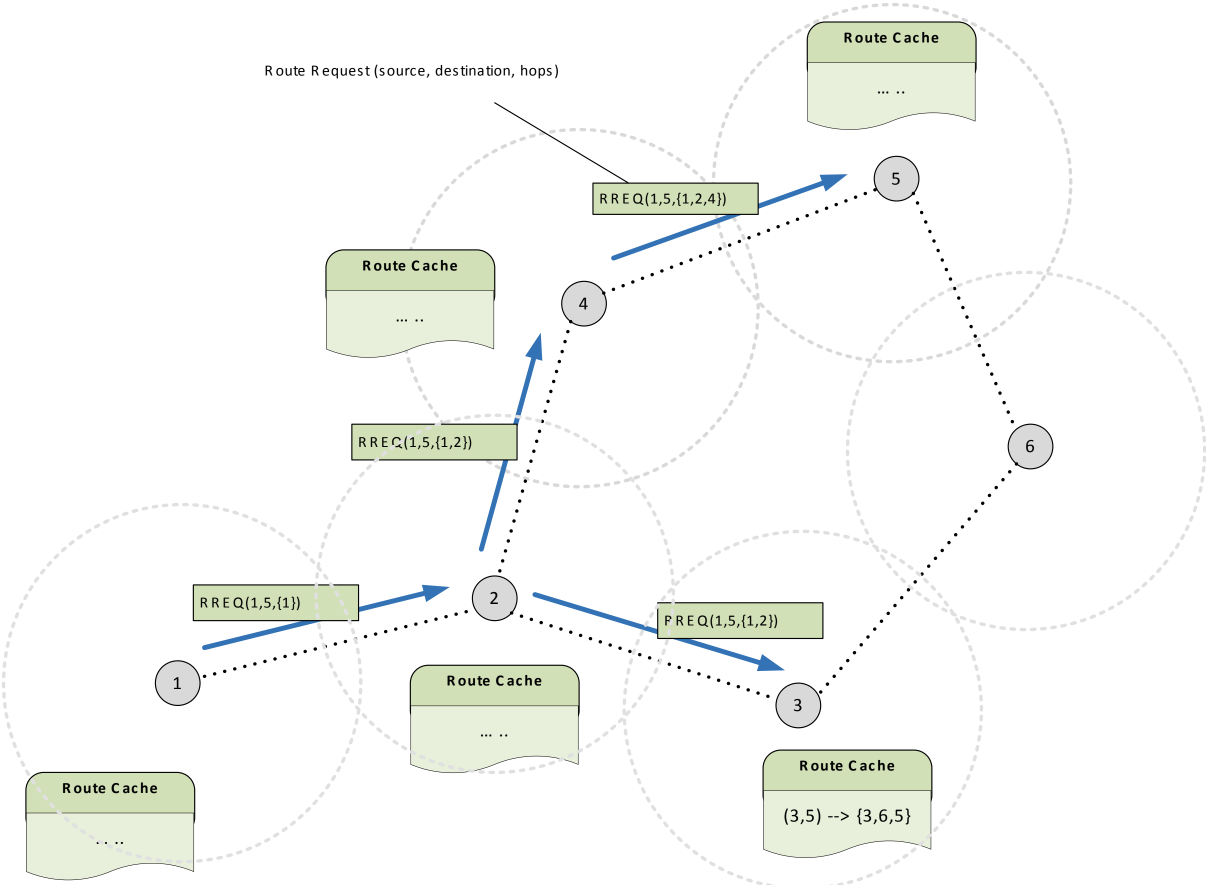

Figure 7-5. Route discovery and Route Cache concept – Route Request phase.

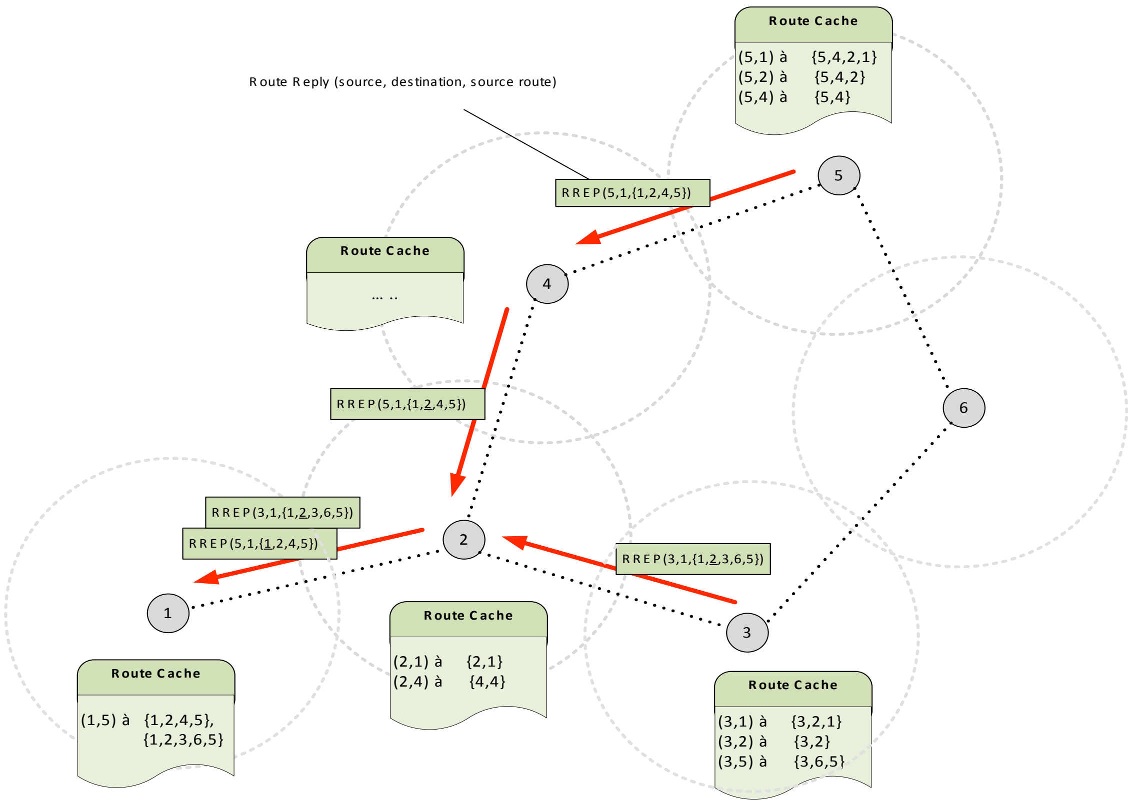

Figure 7-6. Route discovery and Route Cache concept – Route Reply phase.

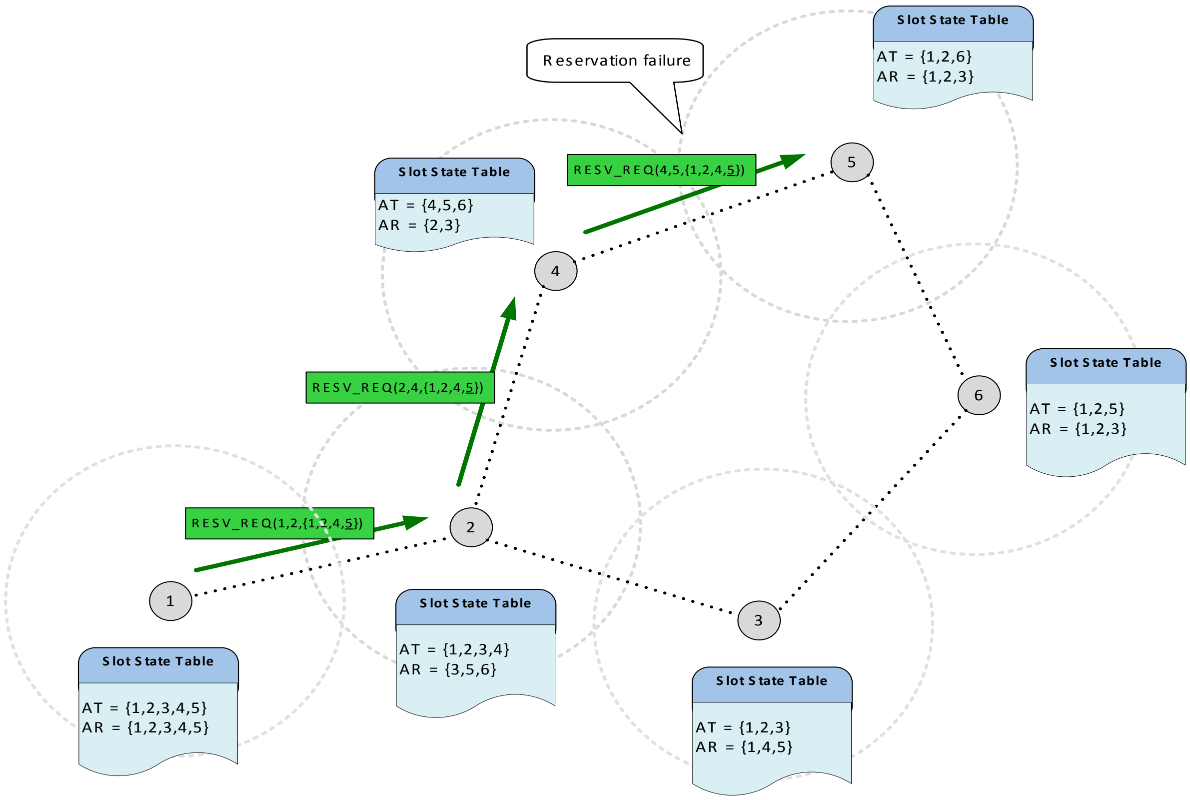

Figure 7-7. End-to-end reservation failure in the presence of multiple paths. AT and AR denote the set of slots available for transmission and reception respectively.

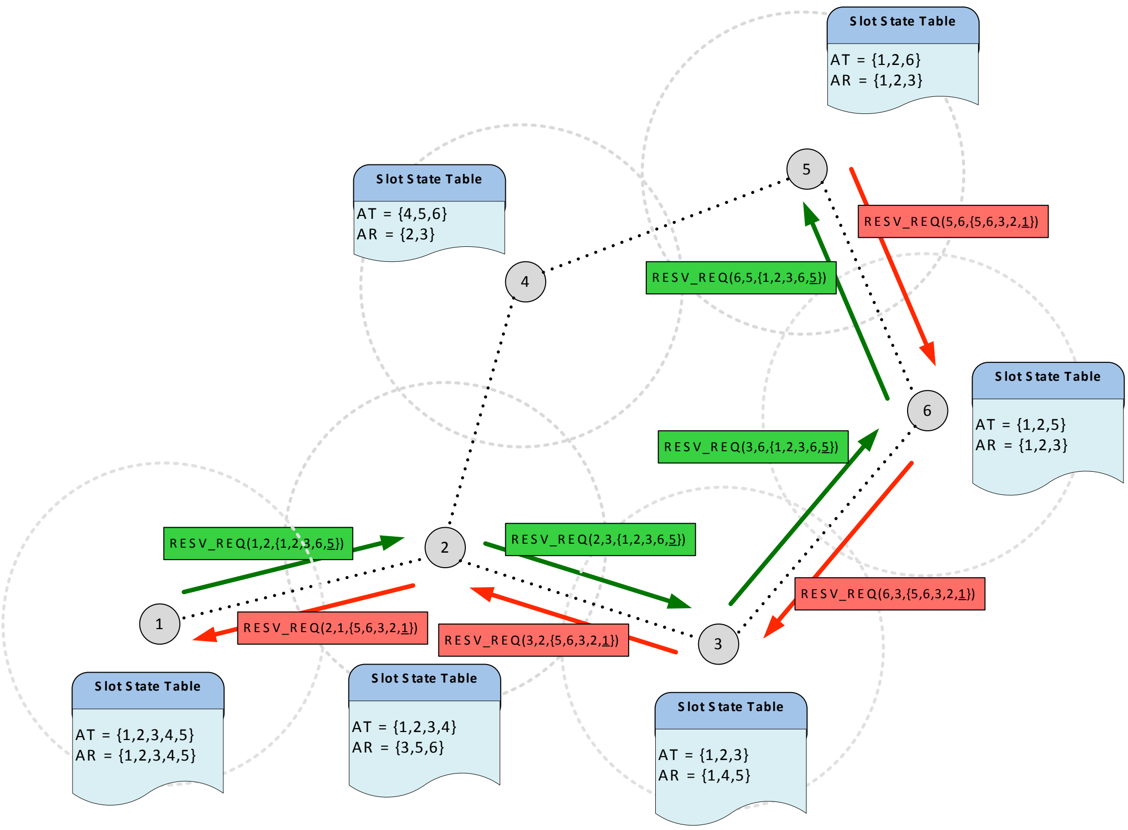

Figure 7-8. End-to-end reservation success in the presence of multiple paths.

7.5 Performance evaluation

In this section, we present simulation results regarding performance evaluation of EERV protocol. The evaluation is performed through a set of tests using the NS-2 simulator. In the simulation experiments, we compare the performance of our protocol with the EDCF combined with DSR routing and DARE scheme for a variety of topologies and scenarios.

7.5.1 Simulation Model

We consider an ad-hoc network composed of 100 nodes randomly distributed on 2×2 km^2 area. The wireless channel has 2 Mbps bit rate. The transmission range of nodes is 250 meters. Each node can be the source of a G.711 voice flow that generates packets of 160 bytes payload at 64 Kbps rate. The talkspurt and silence durations are set to 1.35s and 1s respectively. For data traffic, we used Poisson traffic sessions with average rate of 64 Kbps. Voice sessions and data sessions are started at random instants of the simulation and have duration of 200s.

For EDCF, we use the simulation model of Wiethölter and Hoene [WH 03]. For DARE, we implemented the scheme as described in [CAR 05 and CAR 06]. The contention parameters for EDCF and DARE are shown in Table 7-1.

Table 7-1. Contention parameters for DCF and EDCF

Traffic category	CWmin	CWmax	AIFSN
Background traffic	31	1023	7
Voice traffic	7	15	2
DCF	31	1023	2

7.5.2 Simulation results

7.5.2.1 Impact of data traffic load

In this study, we analyze the impact of traffic load on the performance of EERV. We consider that the 100 nodes are not mobile.

Figure 7-9 shows the frequency of reservation breakage depending on the number of hops for different traffic loads (10, 20, and 30 voice connections). With EERV, the reservation breakage frequency remains null with the three considered traffic loads, while the traffic load has a huge impact on the reservation breakage with DARE. With 10 connections, a voice reservation breaks every 50s in average. A voice connection breaks every 25s with 20 connections, and every 10s with 30 connections.

As the traffic load increases, the number of reservation breakage increases because of the higher probability that reservation control packets collide at neighbor nodes. As these neighbors do not record reservations, they cause interference during reserved slots when they try to reserve already reserved slots. EERV reduces the reservation breakage frequency through providing guarantees that reservation control packets transmitted by a node are received by all its neighbors during the reservation setup phase, which is confirmed by the low number of reservation breakage.

Figure 7-10 shows the average reservation establishment delay with EERV and DARE with the increase of hops. EERV achieves lower establishment delay than DARE. The high establishment delay of DARE compared to EERV is mainly due to the high collision rate of reservation packets with DARE, especially at high data traffic load.

Figure 7-11 shows the voice packet dropping rate achieved by EERV, DARE and EDCF-DSR for different traffic loads. The three protocols achieve low dropping rate with 10 connections. With 20 and 30 connections, EDCF-DSR achieves the highest dropping rate. This high dropping rate is due to the high contention and collision rates with EDCF-DSR compared with EERV and DARE. However, the dropping rate is not impacted by the traffic load with EERV, while it increases with DARE. This is explained by the increase of reservation breakage rate of DARE with the increase of traffic load as illustrated in Figure 7-9.

Figure 7-12 shows the average end-to-end delay of voice traffic achieved by EERV, DARE, and EDCF-DSR with the increase of hops for different traffic loads. As the traffic load increases, nodes with EDCF-DSR experience more contentions, and thus more collisions and wider backoff, the fact which explains the higher delay of EDCF-DSR. We see that the traffic load has an important effect on the voice delay with DARE, and not with EERV.

Figure 7-13 shows the acceptance ratio of end-to-end reservations with EERV and DARE with the increase of the number of hops for different traffic loads. At a traffic load of 10 connections, both protocols achieve a reservation acceptance of 100%. As the number of connections and the number of hops increases, the reservation acceptance ratio decreases due to the lack of bandwidth availability. However, EERV achieves higher reservation acceptance ratio than DARE. The low reservation acceptance ratio of DARE is explained by its high reservation failure at high number of connections due to bandwidth non-availability. As DARE tries to reserve bandwidth only on the shortest path, end-to-end reservation requests are rejected if the shortest path does not provide the required bandwidth. Contrarily to DARE, EERV is tries to reserve bandwidth on several paths. Consequently, the chance of end-to-end reservation failure with EERV is lower than with DARE.

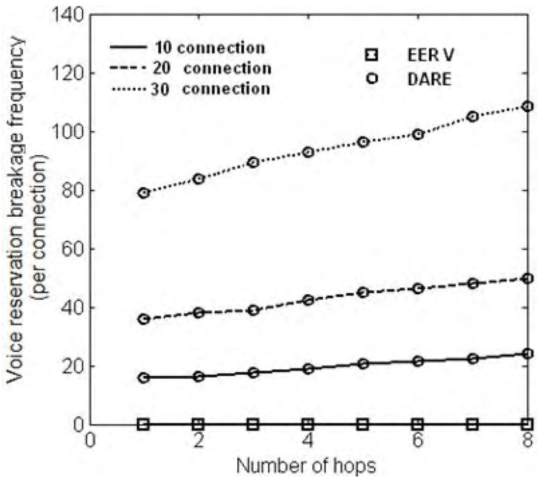

Figure 7-9. Voice reservation breakage frequency with the increase of hops.

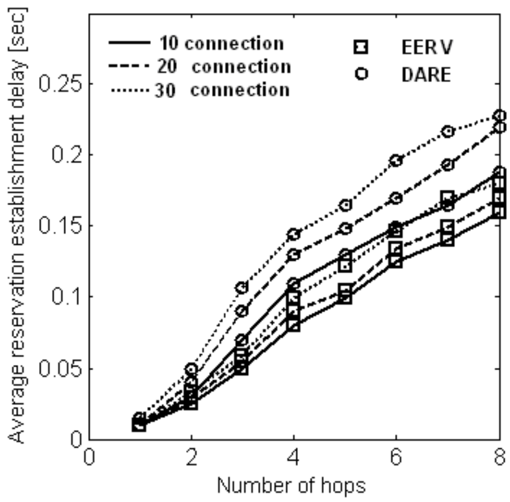

Figure 7-10. Average reservation establishment delay with the increase of hops.

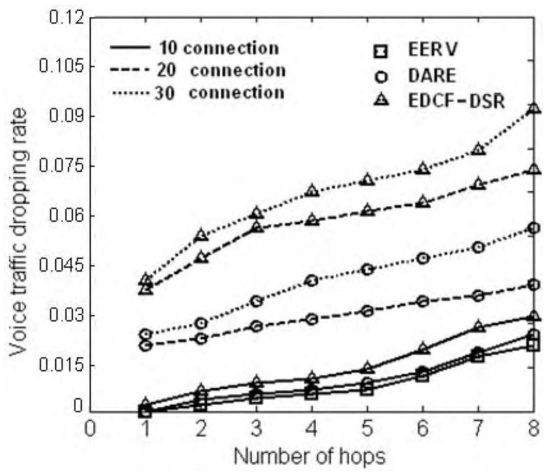

Figure 7-11. Voice packet dropping rate with the increase of hops.

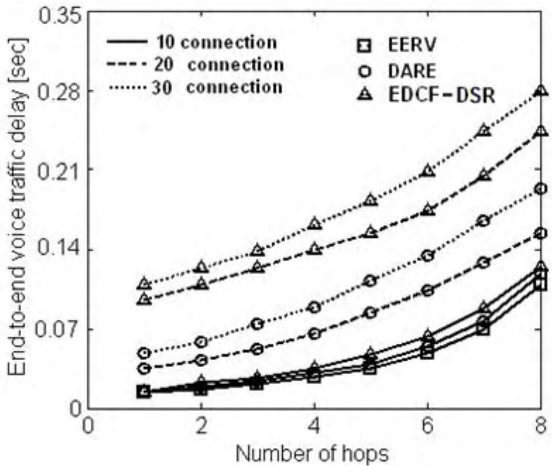

Figure 7-12. Voice traffic delay with the increase of hops.

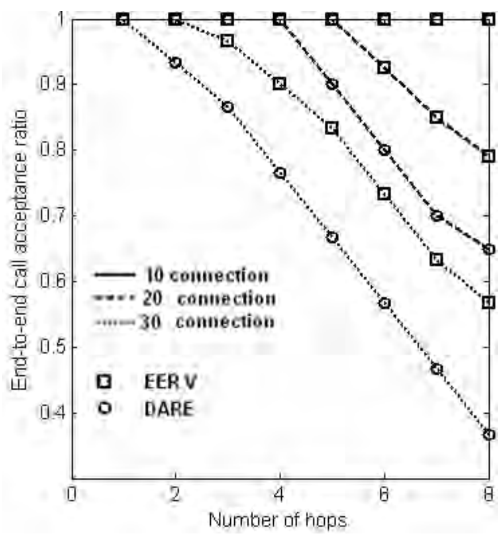

Figure 7-13. End-to-end reservation acceptance ratio with the increase of the number of hops.

7.5.2.2 Impact of mobility

In this study, we analyze the effect of mobility on the performance of EERV in terms reservation breakage frequency, end-to-end delay, and dropping rate. For mobility, we use a modified RWP (Random Way-Point) model. Each node chooses randomly its next position and moves toward that position. The node stays in its new position for a pause time *dt* after which it chooses another position. In order to avoid the well-known problem of low average nodal speed with RWP in the long run [YLN 03], the speed is not chosen randomly between 0 and a maximal speed V_{max}, but set to the desired speed.

We fix the traffic load to 20 voice and 20 data connections, and increase the mobility from 2 m/s to 15 m/s with an increment of 2 m/s.

Figure 7-14 shows the reservation breakage frequency of voice connections achieved by DARE and EERV with the increase of mobility with different pause times. The figure shows that both of DARE and EERV are affected by mobility. Reservation breakage increases as mobility increases and as pause time decreases. However, with EERV reservation breakage occurs less frequently compared to DARE thanks to its efficient reservation scheme at the MAC sub-layer.

Figure 7-15 and Figure 7-16 show the end-to-end voice traffic delay and voice packet dropping rate achieved with EERV, DARE, and EDCF-DSR respectively with the increase of mobility. At pause time of 10 and 20s, EERV and DARE outperform EDCF-DSR. This is because at high pause times the network topology is more stable and reservation disruptions occur less frequently. However, EERV achieves lower dropping rate than DARE. For pause time of 5 s, we see that EDCF outperforms EERV and DARE. At high speed, the dropping rate of DARE and EERV increases up to 15 % (compared to 12 % with EDCF-DSR), and the delay increases to 300 ms compared to 250 ms with EDCF-DSR. These low performances are because the benefit gained from reservations and contention-free transmission decreases. At high mobility and low pause times, the topology changes very frequently, and nodes spend the most of the time moving leading to an increase of reservation disruption. As a consequence of frequent reservation disruption, voice packets are queued longer and dropped more frequently.

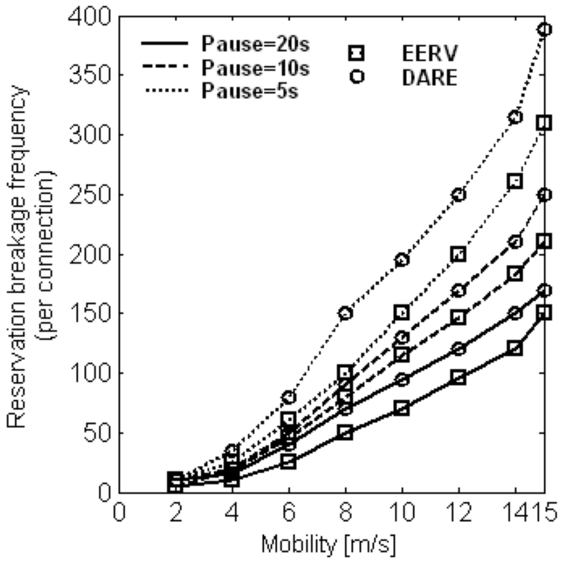

Figure 7-14. Voice connections breakage frequency with the increase of mobility.

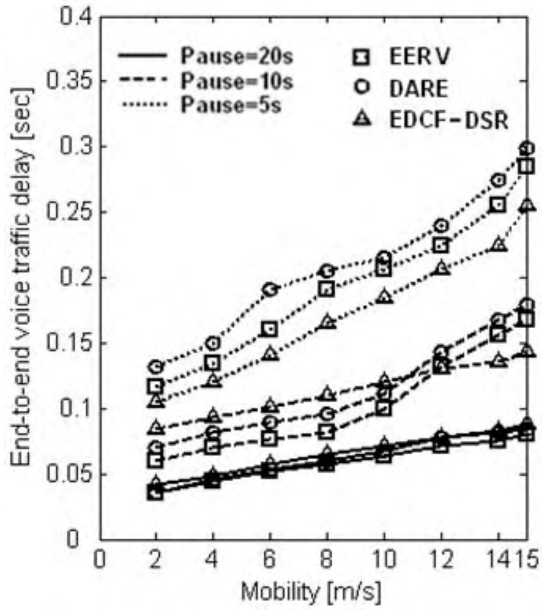

Figure 7-15. Voice traffic delay with the increase of mobility.

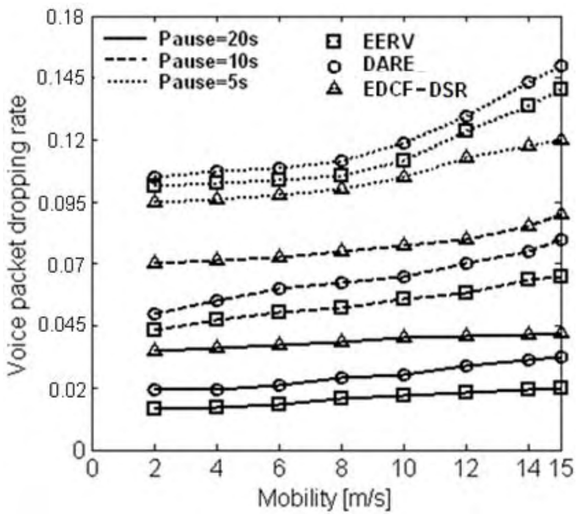

Figure 7-16. Voice packet dropping rate with the increase of mobility.

7.6 Conclusion

In this chapter, we present an efficient end-to-end bandwidth reservation scheme for wireless multi-hop ad-hoc networks. We extend the reservation scheme presented in chapter 5 in order to enable end-to-end reservation and deal with performance degradation due to mobility.

For end-to-end reservation, we propose extensions to be made at the network layer in the form of modules that control reservation and releasing of bandwidth along a path in cooperation with the MAC sub-layer. The particular feature of our end-to-end reservation scheme is that it does not try to reserve bandwidth only on the shortest path. Instead, it considers other alternative paths (if available) if the shortest path does not provide the required bandwidth. This enables lowering the end-to-end reservation failure rate.

Depending on node mobility pattern, mobility may cause frequent reservations breakage resulting in significant performance degradation. In order to reduce the negative impact of mobility, we propose a reservation breakage detection and reservation recovery mechanisms.

Experiments show that EERV still performs better, although the performance gains are slightly reduced due to mobility of nodes. Therefore, EERV is an efficient solution for bandwidth reservation for voice traffic support over ad-hoc networks with low path breakage frequency such as networks with low-mobility nodes or networks where nodes move following a group pattern.

Chapter 8: Conclusion and Perspectives

Wireless communications is one of the fastest growing technologies in the word. With the growing of this technology, the need of mobility of end-users does not stop increasing. Mobile ad-hoc networks are interesting for end-users as they allow flexible mobility. These networks must be able to support the same applications as wired networks, in particular multimedia and real-time applications. These applications are characterized by strict QoS requirements such as minimum guaranteed bandwidth, maximum transmission delay, and maximum dropping rate. A key feature in fulfilling these QoS requirements is the efficient utilization of the limited bandwidth. Thus, QoS support can be delivered to the end-users only if the radio channel is used efficiently.

In this thesis, we focus on QoS support in wireless ad-hoc networks. To address this issue, we decided to operate at the data link layer, where the medium access control protocol plays an important role in the sharing and utilization of bandwidth. Thus, mobile ad-hoc networks can provide the required QoS only if MAC protocols grant channel to nodes in such a manner that increases bandwidth utilization, and reduces channel access delay. In fact, these two parameters are strongly affected by the packet collision rate at the MAC sub-layer, and the objective of achieving efficient bandwidth utilization and low access delay can be fulfilled only through keeping the collision rate at low level.

Our contributions consist in proposing a bandwidth reservation scheme to provide QoS guarantees in wireless ad-hoc networks. The main aims of this scheme are to reduce the collision rate, and achieve efficient bandwidth utilization.

8.1 Synthesis

In the first part of our work we proposed a reservation-based MAC protocol for wireless ad-hoc networks. The proposed protocol aims at achieving efficient bandwidth utilization through reducing the packet collision rate. The protocol resolves a problem of paramount importance in reservation-based MAC protocols, which is reservation inconsistency. This problem occurs when some conflicts of reservations appear between neighbor nodes, because some of these nodes are not aware of reservations established by their neighbors. Our protocol tries to avoid such conflicts through better coordination and cooperation between nodes. Indeed, the handshake scheme of our protocol consists in ensuring that a reservation is confirmed and considered only if it is recorded by all the neighbors of both the sender and receiver. Any unheard reservation due to collision of reservation control packets during the reservation phase is considered invalid.

In the second part of this thesis, we propose ARPV, which is an adaptation of the proposed reservation scheme in order to take into account the characteristics of voice traffic, and allow its integration with data traffic. The protocol consists in a reservation protocol that controls medium access of both voice and data sources. Due to the different characteristics of voice codecs, several parametrization issues rise in the design of ARPV. The first one of these issues is the choice of the data-slot length that provides the best performance. We compared and analyzed three solutions for this issue: long slot length, small slot length with fragmentation, and small slot length without fragmentation. We found that a choice of short slot length is suitable for codecs with small packets, but incurs a significant waste of bandwidth for codecs with long packets due to fragmentation overhead. A choice of long slot length avoids excessive fragmentation overhead, but results in under-utilization of data-slots when codecs with short packets are used. However, the best solution consists in choosing short slots and allowing voice sources with long packets to reserve contiguous small slots.

The second issue is contention resolution during the reservation phase. We proposed two contention resolution schemes that control the permission probability of voice and data sources in sending their reservation requests. In the first scheme, that we called static priority contention scheme, voice and data sources use fixed permission probabilities with higher priority for voice traffic sources. In the second scheme, that we called the dynamic priority contention scheme, the permission probability of voice and data sources is adapted to the traffic load in the network. We found that the dynamic priority contention scheme gives better performance for both voice and data traffic.

The third issue is the manner data and voice sources send their packets during the data sub-frame. For voice traffic sources, we showed that reserving bandwidth for voice sources for all the duration of the connection leads to a significant waste of bandwidth due to silence periods. In order to reduce this waste of bandwidth, we define multiplexing mechanisms where data traffic sources are enabled to share bandwidth with voice traffic sources so that bandwidth is used efficiently. Thus, we proposed to make voice sources release temporarily their reserved slots when they go to the sleep mode, so that data sources can use them for transmission through either contention (in the Contend for Each Packet scheme), or through reserving these slots temporarily (in the Reserve Temporarily Released slots scheme).

We analyzed the performance of our solution for bandwidth reservation for voice traffic support over MANETs through a stochastic model as well as simulations. The stochastic model and simulations allowed us to dimension efficiently our solution. Thus, they allowed us to analyse the impact of the protocol parameters such as the permission probabilities, to compare static and dynamic priority contention schemes, and to compare the behavior of ARPV with the RTR and CEP schemes. The results show that the RTR scheme is more efficient than the CEP scheme for both voice and data traffic, especially when combined with the dynamic contention scheme. In comparison with the existing standards, results show that our solution is more efficient than the EDCF scheme.

The third part of the thesis was devoted to end-to-end bandwidth reservation in multi-hop ad-hoc networks. We proposed a reservation scheme called End-to-End Reservation scheme for Voice and data traffic support (EERV) which is an extension of ARPV to reserve resources along a path in cooperation with the routing layer. The particular feature of EERV scheme is that it does not try to reserve bandwidth only on the shortest path. Instead, it explores other alternative paths if the shortest path does not provide the required bandwidth. In addition to end-to-end bandwidth reservation, EERV includes mechanisms that alleviate performance degradation due to mobility of nodes.

8.2 Future directions

While the contributions made in the context of this thesis have solved some problems, some features related to our work remain open research fields. These features can be considered as promising future directions for the continuity of our work. Some of the future directions that can be investigated are:

- *Bandwidth reservation in wireless mesh networks:* As pointed out in chapter 7, EERV is an efficient solution for bandwidth reservation in ad-hoc networks with low path breakage frequency such as networks with low-mobility nodes. Thus, the proposed reservation scheme could be useful to provide QoS in wireless mesh networks. However, the traffic load in wireless mesh networks backbone is more important than in wireless ad-hoc networks. Thus, we plan to evaluate the performance of our solution in a wireless mesh network backbone, and adapt it to the characteristics of wireless mesh networks.

- *Bandwidth reservation for video traffic support:* As mentioned, the proposed reservation protocol can be used to provide guaranteed access to different real-time traffic types. In this thesis, we adapted this protocol to take into account the characteristics of voice traffic. However, it is interesting to explore extensions that can be done in order to support VBR-video traffic. As VBR-video sources have variable bandwidth requirements, the future protocol should allocate bandwidth to these sources based on their actual and future bandwidth requirements. Thus, first of all, bandwidth estimation scheme that estimates the future required bandwidth is needed in order to perform reservations in advance, and avoid buffer overflow during excessive bursts.

- *Multi-path bandwidth reservation for video:* As video sources are bandwidth demanding, end-to-end resource reservation for video traffic represents a serious problem. In high traffic load conditions it is difficult to find a path that satisfies the required bandwidth. Thus, we plan to extend EERV by including Multipath reservation. The future protocol tries to reserve bandwidth on one of the available paths, and makes several partial reservations on different paths if it is not possible to reserve the required bandwidth on a single path.

- *Multi-channel bandwidth reservation:* As defined in the IEEE 802.11 standard, the physical layer supports the use of multiple radio channels, leading to an increase of the available bandwidth. In current solutions for multi-channel access, time is divided into beacon intervals. An amount of each interval is used to negotiate channels for transmission during the current beacon interval. Thus, to reserve one of the channels for transmission during the current beacon interval, each node follows a three-way handshake messaging to negotiate the channel that is going to be used for transmission with the intended receiver. Once the corresponding channel is agreed, the sender and receiver switch to agreed-upon channel and try to exchange a packet through the conventional RTS/CTS/DATA/ACK handshake. One of the drawbacks of such scheme is the sender and receiver are required to use the backoff scheme twice: before transmitting channel negotiation messages, and before the RTS/CTS/DATA/ACK handshake. However, while the use of multiple channels offers more bandwidth, current solutions for multi-channel utilization are not suitable for delay-sensitive traffic due to effect of contention. Thus, one of the promising research directions is to extend our reservation MAC protocol in order to take into account the advantage of the use of multiple channels while satisfying the delay requirements.

Publications

Book chapters

[1] G. Boudour, Mahboub Bali, Cédric Teyssié, "MAC protocols for Quality of Service provisioning in mobile ad hoc networks". In End-to-End Quality of Service Engineering in Next Generation Heterogenous Networks, Wiley, pp. 213-247, November 2008.

International Journals

[2] G. Boudour, C. Teyssié, Z. Mammeri, "Reservation MAC Protocols for Ad-Hoc Networks: Analysis of the approaches". In International Journal of Business Data Communication and Networking, IGI Global, Volume 5, No 2, pp.52-67, April 2009.

[3] G. Boudour, Z. Mammeri, C. Teyssié, "End-to-End Bandwidth Allocation Scheme for Voice Traffic Support over MANETs". In Telecommunication Systems Journal, Springer, Volume 48, No 3, November 2011 (to appear).

International conferences

[4] G. Boudour, C. Teyssié, Z. Mammeri, "Adaptive Reservation MAC Protocol for Voice Traffic in Wireless Ad-hoc Networks". In The Second IFIP Wireless and Mobile Networking Conference (WMNC'09), Gdansk, Poland, September 2009.

[5] G. Boudour, C. Teyssié, Z. Mammeri, "Resource Reservation for Voice Support in MANETs". In The Fifth International Conference on Wireless Communications, Networking and Mobile Computing (WICOM'09), Beijing, China, September 2009.

[6] G. Boudour, C. Teyssié, Z. Mammeri, "Robust Reservation Protocol for Wireless Ad-hoc Networks". In The 5th ACM International Wireless Communications and Mobile Computing Conference (IWCMC'09), Leipzig, Germany, June 2009.

[7] G. Boudour, C. Teyssié, Z. Mammeri, "Bandwidth Reservation for Heterogeneous Traffic Classes in Mobile Wireless Networks". In The Fifth Advanced International Conference on Telecommunications (AICT'09), Venice, Italy, May 2009.

[8] G. Boudour, C. Teyssié, Z. Mammeri, "A Robust Reservation Protocol For Wireless Ad-hoc Networks". In The First International Conference on Communication Systems and Networks (COMSNETS'09), Bangalore, India, January 2009.

[9] G. Boudour, C. Teyssié, Z. Mammeri, "Reservation clash handling to optimize Bandwidth Utilization in MANETs". In International conference on Communication Theory, Reliability and Quality of Service (CTRQ'08), Bucharest, Romania, July 2008.

[10] G. Boudour, C. Teyssié, Z. Mammeri, "Scheduling-based Reservation MAC Protocol for Bandwidth and Delay Optimization in Wireless Mesh Networks". In IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob'08), Avignon, France, October 2008.

[11] G. Boudour, C. Teyssié, Z. Mammeri, “Performance Analysis of Reservation Protocols for Ad-Hoc Networks”. In IFIP International Conference on Mobile and Wireless Communications Networks (MWCN’08), Toulouse, France, October 2008.

Bibliography

[802_15] IEEE 802.15. http://grouper.ieee.org/groups/802/15/

[802_16] IEEE Working Group 802.16 on Broadband Wireless Access Standards. http://grouper.ieee.org/ groups/802/16/

[ACC] "La Qualité de Service le la Voix sur IP, Principes et Assurance", *www.accellent-group.com*.

[AND 04] S. Andersen, et al., "Internet Low Bit Rate Codec (iLBC)", RFC 3951, December 2004.

[BAH 06] M. Bhatia, et al., "Voice over IP Fundamentals", Second Edition, October 2006, Cisco-Press.

[BDM 01] R. O. Baldwin, N .J. Davis IV, S. F. Midkiff, "Packetized Voice Transmission Using RT-MAC, a Wireless Real-time Medium Access Control Protocol", *ACM SIGMOBILE Mobile Computing and Communications Review*, Vol. 5, N° 3, pp. 11-25, July 2001.

[BDM 99] R. O. Baldwin, N .J. Davis IV, S. F. Midkiff, "A Real-time Medium Access Control Protocol for Ad-Hoc Wireless Local Area Networks", *ACM SIGMOBILE Mobile Computing and Communications Review*, Vol. 3, N° 2, pp. 20-27, April 1999.

[BLUE] Bluetooth Special Interest Group. http://www.bluetooth.org/

[BMS 04] T. Bheemarjuna Reddy, B. S. Manoj, and C. Siva Ram Murthy, "Multimedia Traffic Support for Asynchronous Ad hoc Wireless Networks", *IEEE BROADNETS 2004*, pp. 569-578, October 2004.

[BOU 06] A. Boukerche, "Handbook of Algorithms for Wireless Networking and Mobile Computing", *Chapman & Hall/CRC Computer and Information Science Series*, 2006.

[BTM 08-1] G. Boudour, C. Teyssié, Z. Mammeri, "Performance Analysis of Reservation MAC protocols for Ad-Hoc Networks", *IFIP International Conference on Mobile and Wireless Communications Networks (MWCN 2008)*, Toulouse, pp. 173-186, September 2008.

[BTM 08-2] G. Boudour, C. Teyssié, Z. Mammeri, "*Reservation clash handling to optimize bandwidth utilization in MANETs*", *International Conference on Communication Theory, Reliability, and Quality of Service (CTRQ 2008),* Bucharest, Romania, pp. 77-82, June 2008.

[CAR 04] E. Carson, et al., "Distributed Allocation of Time Slots for Real-time Traffic in a Wireless Multi-hop Network", *European Wireless*, Barcelona, Spain, February 2004.

[CAR 05] E. Carson, et al., "A Performance Comparison of QoS Approaches for Ad Hoc Networks: 802.11e versus Distributed Resource Allocation", *European Wireless*. Nicosia, Cyprus, April 2005.

[CAR 06] E. Carson et al., "A Distributed End-to-End Reservation Protocol for IEEE 802.11-Based Wireless Mesh Networks", *IEEE Journal on Selected Areas in Communications*, Vol. 24, No.11, November 2006.

[CHE 02] Y. Chen, et al., “On-demand, Link State Multi-path QoS Routing in a Wireless Mobile Ad-hoc Networks”, *European Wireless’02*, pp. 135-141, February 2002.

[CIS 06] “Understanding Jitter in Packet Voice Networks”, Cisco IOS Platforms, Document ID: 18902, February 2006.

[CIS 06-2] “Understanding Delay in Packet Voice Networks”, Cisco IOS Platforms, Document ID: 5125, February 2006.

[CKC 03] W. A. Chang, C. G. Kang, Y. Z. Cho, “Soft Reservation Multiple Access with Priority Assignment (SRMA/PA): A Distributed MAC Protocol QoS-guaranteed Integrated Services in Mobile Ad-hoc Networks”, *IEICE Transactions on Communications*, Vol. E86-B, N° 1, pp. 50-59, January 2003.

[CN 99] S. Chen, K. Nahrstedt, “Distributed Quality-of-Service Routing in Ad-hoc Networks”, *IEEE Journal on Selected Areas in Communications*, Vol. 17, N° 8, pp. 1488–1504, August1999.

[DC 99] D. J. Deng, R. S. Chang, “A Priority Scheme for IEEE 802.11 DCF Access Method”, *IEICE Transactions Communications*, Vol. E82-B(1), pp. 96-102, January 1999.

[DE 02] S. De, et al., “Trigger-based Distributed QoS Routing in Mobile Ad hoc Networks”, *ACM SIGMOBILE Mobile Computing and Communications Review*, Vol. 6, N° 3, pp. 22-35, July 2002.

[DHA 02] D. Dharmaraju, et al., “INORA - A Unified Signaling and Routing Mechanism for QoS Support in Mobile Ad hoc Networks”, *International Conference on Parallel Processing Workshops (ICPPW’02)*, pp. 86-93, August 2002.

[ESP 08] David Espes, “Protocoles de routage réactifs pour l’optimisation de bande passante et la garantie de délai dans les réseaux ad hoc mobiles”, PhD thesis, Paul Sabatier University, Toulouse, November 2008.

[GAH 02] A. Gahng-Seop, et al., “Supporting Service Differentiation for Real-Time and Best-Effort Traffic in Stateless Wireless Ad Hoc Networks (SWAN)”, *IEEE Transactions on Mobile Computing*, Vol. 1, N° 3, pp. 192-207, July 2002.

[GOD 88] D. J. Goodman et al., “Packet Reservation Multiple Access for Local Wireless Communications”, *IEEE Vehicular Technology Conference*, Philadelphia, pp. 701-706, June 1988.

[HAR 03] W. C. Hardy, “VoIP Service Quality, Measuring and Evaluating Packet-Switched Voice”, 2003, McGraw-Hill.

[HOMERF] HomeRFWorking Group. http://www.consortiuminfo.org/links/homerf.php

[IEE 05] IEEE 802.11WG. Draft Supplement to Standard for Telecommunications and Information Exchange between Systems-LAN/MAN Specific Requirements-part11: MAC Enhancements for Quality of Service (QoS). IEEE 802.11e Standard Draft/D13.0, 2005.

[IETF] The Internet Engineering Task Force (IETF). http://www.ietf.org/

[INW 04] J. Inwhee, “Qos-Aware MAC with Reservation For Mobile Ad-Hoc Networks”, *IEEE Vehicular*

Technology Conference, September 2004.

[ITU 03] ITU Rec. G.131, "G.131: Talker echo and its control", International Telecommunication Union, November 2003.

[ITU 06] ITU Rec. G.723.1, "G.723.1: Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s", International Telecommunications Union, May 2006.

[ITU 07] ITU Rec. G.729, "Coding of speech at 8 kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP)", International Telecommunications Union, January 2007.

[ITU 88] ITU Rec. G.722, "G.722: 7 kHz audio-coding within 64 kbit/s", International Telecommunications Union, November 1988.

[ITU 90] ITU Rec. G.726, "G.726: 40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)", International Telecommunications Union, December 1990.

[ITU 92] ITU Rec. G.728, "G.728: Coding of Speech at 16 kbit/s Using Low-Delay Code Excited Linear Prediction", International Telecommunications Union, September 1992.

[ITU 93] ITU Rec. G.711, "G.711: Pulse Code Modulation (PCM) of Voice Frequencies," International Telecommunications Union, February 1993.

[ITU 96] ITU Rec. G.114, "One-way Transmission Time", International Telecommunications Union, February 1996.

[ITU 96-2] ITU Rec. G.114, "One-Way Transmission Time," International Telecommunication Union, 1996.

[JL 98] J. Ju and V. O. K. Li, "An Optimal Topology-Transmission Scheduling Method in Multihop Packet Radio Networks", *IEEE/ACM Transactions on Networking*, Vol. 6, No. 3, pp. 298-306, June 1998.

[JL 99] J. Ju and V. O. K. Li, "TDMA Scheduling Design of Multihop Packet Radio Networks Based on Latin Squares", *IEEE Journal on Selected Areas in Communications*, Vol. 17, No. 8 1999, pp. 1345-1352, August.

[JMB 01] B. D. Johnson, D. A. Maltz, J. Broch, "DSR: The Dynamic Source Routing Protocol for Mobile Ad hoc Networks" In C. E. Perkins (Ed.), *Ad Hoc Networking*, Addison Wesley, pp. 139-172, 2001.

[JMB 96] D. B. Johnson, D. A. Maltz, J. Broch, "Dynamic Source Routing in Ad Hoc Wireless Networks" In T. Imielinski and H. Korth (Ed.), *Mobile Computing*, Kluwer Academic Publishers, pp. 153–181, 1996.

[JOE 04] I. Joe, "Qos-Aware MAC with Reservation for Mobile Ad-Hoc Networks", *IEEE Vehicular Technology Conference,* Los Angeles, pp. 1108-1112, September 2004.

[KAN 02-1] V. Kanodia, et al., "Distributed priority scheduling and medium access in ad hoc networks", *ACM/Baltzer journal of Wireless Networks*, Vol. 8, N° 5, pp. 455-466, September 2002.

[KAN 02-2] V. Kanodia, et al., "Ordered Packet Scheduling in Wireless Ad Hoc Networks: Mechanisms and Performance Analysis", *ACM MOBIHOC 2002*, pp. 58-70, June 2002.

[KPP 05] P. Koutsakis, S. Psychis, and M. Paterakis, “Integrated Wireless Access for Videoconference From MPEG-4 and H.263 Video Coders With Voice, E-mail, and Web Traffic”, *IEEE Transactions On Vehicular Technology,* Vol. 54, N° 5, 2005.

[KUN 05] I.P. Kun, “QoS In Packet Networks”, Springer, 2005.

[LG 97] C. H. R. Lin and M. Gerla, “Asynchronous Multimedia Multi Hop Radio Network”, *IEEE INFOCOM’97*, Kobe, Japan, pp. 118-125, April 1997.

[MG 06] X. Meng, J. Garcia-Luna-Aceves, “Channel Access Using Opportunistic Reservation in Ad-Hoc Networks”, *International Conference on Mobile Ad-hoc and Sensor Systems (MASS 2006),* Vancouver, Canada, pp. 71-80, October 2006.

[MIK 88] V. A. Mikhailov, “Geometrical analysis of the stability of Markov chains in R^n_+ and its application to throughput evaluation of the adaptive random multiple access algorithm”, *Problemy Peredachi Informatsii***,** Vol. 24, No. 1, pp. 61-73, July 1988 *(translated in Problems of Information Transmission, July 1988).*

[MMS 07] J. V. Meggelen, L. Madsen, and J. Smith, “Asterisk™: The Future of Telephony”, Second edition, August 2007, O’REILLY.

[MS 02] B. S. Manoj, C. Siva Ram Murthy, “Real-time Traffic Support For Ad hoc Wireless Networks”, *IEEE ICON 2002*, pp. 335-340, August 2002.

[MVS 04] B. S. Manoj, V. Vidhyashankar, and C. Siva Ram Murthy, “Slot Allocation Strategies for Delay Sensitive Traffic Support in Asynchronous Ad hoc Wireless Networks”, *Wireless Communications and Mobile Computing Journal*, Vol. 5, N° 2, pp. 193-208, August 2004.

[NS2] The Network Simulator NS-2 (http://www.isi.edu/nsnam/ns/)

[PAL 01] Pal A., “Distributed MAC Layer Protocols for Real-Time Communication in Ad-Hoc Wireless Networks”, M.S. Thesis, Ohio State University, 2001.

[PDO 02] A. Pal, A. Dogan, F. Ozguner, “MAC layer protocols for real-traffic in ad-hoc networks”, *IEEE International Conference Parallel Processing*, 2002.

[PRD 00-1] C. E. Perkins, E. M. Royer, and S. R. Das, “Quality of Service in Ad hoc On-Demand Distance Vector Routing”, IETF Internet Draft, draft-ietfmanet-qos-00.txt, July 2000 (Work in Progress).

[PRD 00-2] C. E. Perkins, E. M. Royer, S. R. Das, “Quality of Service for Ad hoc On-demand Distance Vector Routing”, IETF Internet Draft, draft-ietf-manet-aodvqos-00.txt, July 2000.

[PW 94] C. E. Perkins, T. J. Watson, “Highly dynamic destination sequenced distance vector routing (DSDV) for mobile computers”, *ACM SIGCOMM’94 Conference on Communications Architectures*, London, UK, 1994.

[RFC 1633] R. Braden, D. Clark, and S. Shenker, “Integrated Services in the Internet Architecture: an Overview”, *IETF RFC 1633*, June 1994.

[RFC 2205] R. Braden, et al., “Resource reservation protocol (RSVP) - Version 1 Functional Specification”, *IETF RFC 2205,* September 1997.

[RFC 2475] S. Blake, et al., “An Architecture for Differentiated Services”, *IETF RFC 2475*, December 1998.

[RFC 3561] C. Perkins, E. Belding-Royer, and S. Das, RFC 3561, “Ad hoc On-Demand Distance Vector

(AODV) Routing”, July 2003.

[RFC 3561] C. Perkins, E. Belding-Royer, and S. Das, “Ad hoc On-Demand Distance Vector (AODV) Routing”, RFC 3561, July 2003.

[RFC 3626] T. Clausen, P. Jacquet, “Optimized Link State Routing Protocol (OLSR)”, RFC 3626, 2003.

[RFC 4728] D. Johnson, Y. Hu, and D. Maltz, RFC 4728, “The Dynamic Source Routing Protocol (DSR) for Mobile Ad Hoc Networks for IPv4”, February 2007.

[RNT 03] L. Romdhami, Q. Ni, and T. Turletti, “Adaptive EDCF: Enhanced Service differentiation for IEEE 802.11 Wireless Ad-Hoc Networks”. *IEEE WCNC conference*, New Orleans, USA, pp. 16-20, March 2003.

[RS 92] I. Rubin, and S. Shamabayati, “Performance evaluation of a reservation random access scheme for packetized wireless systems with call control”, *IEEE GLOBECOM’92*, Orlando, pp. 16-20, December 1992.

[RWO 95] S. Redl, M. Weber, and M. Oliphant, “An Introduction to GSM”, Artech House, 1995.

[SC 03] H. Schulzrinne, S. Casner, “RTP Profile for Audio and Video Conferences with Minimal Control”, RFC 3551, July 2003.

[SEO 00] L. Seoung-Bum, et al., “INSIGNIA: An IP-Based Quality of Service Framework for Mobile ad Hoc Networks”, *Journal of Parallel and Distributed Computing*, Vol. 60, N° 4, pp 374-406, April 2000.

[SHE 02] J. Shengming, et al., “A Simple Distributed PRMA For MANETs”, *IEEE Transactions On Vehicular Technology*, Vol. 51, N° 2, pp. 293-305, March 2002.

[SK 96] J. L. Sobrinho, A. S. Krishnakumar, “Real-time Traffic Over the IEEE 802.11 Medium Access Control Layer”, *Bell Labs Technical Journal*, Vol. 1, pp. 172-187, 1996.

[SK 99] J. L. Sobrinho, A. S. Krishnakumar, “Quality-of-Service in Ad Hoc Carrier Sense Multiple Access Wireless Networks”, *IEEE Journal on Selected Areas in Communications*, Vol. 17, N° 8, pp. 1353-1368, August 1999.

[SM 04] C. Siva Ram Murthy, and B. S. Manoj, “Ad Hoc Wireless Networks: Architecture and Protocols”, *Prentice Hall Communication and Emerging Technologies Series*, 2004.

[SN 02] S. H. Shah, K. Nahrstedt, “Predictive Location-based QoS Routing in Mobile Ad hoc Networks”, *IEEE ICC’02*, Vol. 2, pp. 1022-1027, May 2002.

[TBA 01] A. K. Talukdar, B. R. Badrinath, and A. Acharya, “MRSVP: A Resource Reservation Protocol for an Integrated Services Network with Mobile Hosts”, *ACM/Baltzer Wireless Networks Journal*, Vol. 7, N° 1, pp. 5-19, January 2001.

[TCP] Transmission Control Protocol, RFC 793

[UDP] User Datagram Protocol, RFC 768

[VIV 04] V. Vivek, et al., “A Novel Out-of-band Signaling Mechanism for Enhanced Realtime Support in Tactical Ad-hoc Wireless Networks”, *IEEE Real-time and embedded Technology and Applications Symposium (RTAS’04)*, pp. 56-63, May 2004.

[WH 03] S. Wiethölter, C. Hoene, “Design and Verification of an IEEE 802.11e EDCF Simulation Model in NS-2.26", TKN-03-19 Technical Report Series, Technische Universität Berlin, November 2003.

[WIMAX] WiMAX Forum. http://www.wimaxforum.org/

[XIA 00] H. Xiao, et al., "A Flexible Quality of Service Model for Mobile Ad-hoc Networks", *IEEE Vehicular Technology Conference*, Vol. 1, pp. 445–449, May 2000.

[YLN 03] J. Yoon, M. Liu, and B. Noble, "Random Waypoint Considered Harmful", *IEEE INFOCOM'03*, San Francisco, pp. 1312-1321, April 2003.

[ZC 98] C. Zhu, M. S. Corson, "A Five-Phase Reservation Protocol (FPRP) For Mobile Ad hoc Networks", *IEEE INFOCOM'98*, Vol. 1, pp. 322-331, 1998.

[ZG 99] T. Zhenyu, J. J. Garcia-Luna-Aceves, "A Protocol for Topology-Dependent Transmission Scheduling in Wireless Networks", *IEEE Wireless Communications and Networking Conference*, Vol. 3, pp. 1333-1337, September 1999.

[ZIGBEE] ZigBee Alliance. http://www.zigbee.org/