Design of rate-adaptive MAC and medium aware routing protocols for multi-rate, multi-hop wireless networks

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy (Ph.D.) to:

Electronic and Computer Engineering Division,
School of Engineering and Design,
Brunel University,
United Kingdom.

by:
Shahbaz Khan

2008/2009

Abstract

The IEEE 802.11 standard conformant wireless communication stations have multi-rate transmission capability. To achieve greater communication efficiency, multi-rate capable stations use rate-adaptation to select appropriate transmission rate according to variations in the channel quality. The thesis presents two rate-adaptation schemes, each belonging to one of the two classes of rate-adaptation schemes i.e.(1) *the frame-transmission statistics based schemes,* and *(2) Signal-to-Noise Ratio (SNR) based, closed loop schemes.* The SNR-based rate-adaptation scheme, proposed in this thesis uses a novel mechanism of delivering a receiver's feedback to a transmitter; without requiring any modification in the standard frames as suggested by existing research. The *frame-transmission-statistics* based rate adaptation solution uses an on-demand incremental strategy for selecting a rate-selection threshold. This solution is based on a cross-layer communication framework, where the rate-adaptation module uses information to/from the Application layer along with relevant information from the Medium Access Control (MAC) sub-layer. The proposed solutions are highly responsive when compared with existing rate-adaptation schemes; responsiveness is one of the key factors in the design of such protocols. The novel feedback mechanism makes it possible to achieve frame-loss differentiation with just three frames, avoiding the use of Request To Send/ Clear To Send (RTS/CTS) frames and further delays in this process. Performance tests have affirmed that the proposed rate-adaptation schemes are energy efficient; with efficiency up to 19% in specific test scenarios. In terms of throughput and frame loss-differentiation mechanisms, the proposed schemes have shown significantly better performance.

Routing protocols for Mobile Ad-Hoc Networks (MANETs) use broadcast frames during the route discovery process. The 802.11 mandates the use of different transmission rates for broadcast and unicast (data-) frames. In many cases it causes creation of *communication gray zones*, where stations which are marked as '*reachable neighbours'* using the broadcast frames (using lower transmission rate) are not accessible during normal, unicast communication (mainly at a higher rate). Similarly, higher device density, interference and mobility cause variable medium access delays. The IEEE 802.11e introduces four different MAC level queues for four access categories, maintaining service priority within the queues; which implies that frames from a higher priority queue are serviced more frequently than those belonging to lower priority queues. Such an enhancement at the MAC sub-layer introduces uneven queuing delays. Conventional routing protocols are unaware of such MAC specific constraints and as a result these factors are not considered which result in severe performance deterioration. To meet such challenges, the thesis presents a medium aware distance vector (MADV) routing protocol for MANETs. MADV uses MAC and physical layer (PHY) specific information in the route metric and maintains a separate route per-AC-per-destination in its routing tables. The MADV-metric can be incorporated into various routing protocols and its applicability is determined by the possibility of provision of MAC dependent parameters that are used to determine the hop-by-hop MADV-metric values. Simulation tests and comparison with existing MANET protocols demonstrate the effectiveness of incorporating the medium dependent parameters and show that MADV is significantly better in terms of end-to-end delay and throughput.

Acknowledgments

I am grateful to a number of colleagues, my friends and family who supported me in different ways during the Ph.D.

I would like to thank Prof. Hamed Al-Raweshidy who supervised my research and guided me from time to time. He had been like a very good friend and had been supportive every time I faced a problem; I appreciate all his help and support. I would like to thank all my colleagues at the WNCC who had been of tremendous help and support. I would like to acknowledge the encouraging and supportive attitude of all staff at Brunel University for promoting research. All of them have been absolutely professional.

Lately during my stay at Brunel, I had the opportunity to work with Prof. John Sumpter. As the world's leading scientist in his field, his guidance and moral support had a profound influence on me as a researcher and I would like to treasure that experience. I am extremely grateful to John.

I would like to thank my external examiner, Professor David Edwards from Oxford University, internal examiner, Professor Abdul Sadka and the Chairman, Professor Robert Macredie from Brunel University.

I am grateful to the Higher Education Commission of Pakistan and the University of Engineering & Technology, Peshawar for funding my Ph.D. in the U.K.

I specifically like to thank my uncle, M. Afzal Khan, whose support in many ways has been very vital and made a difference in my career for the better.

In the end, I would like to thank my parents, sister and brothers who had been a source of motivation and strength during this course of time.

Table of Contents

List of Figures

List of Tables

List of abbreviations

AARF	Adaptive ARF
AC	Access Category
AIFS	Arbitration IFS
AIFSN	AIFS Number
AODV	Ad hoc On-demand Distance Vector
AP	Access Point
ARF	Auto-Rate Fallback
BACK	Block ACK
BEB	Binary Exponential Backoff
BSS	Basic Service Set
CAP	Controlled Access Period
CLRA	Cross Layer Rate Adaptation
CS	Carrier Sensing
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
CTS	Clear To Send
CW	Contention Window
DCF	Distributed Coordination Function
DIFS	DCF IFS
DREAM	Distance Routing Effect Algorithm for Mobility
DS	Distribution System
DSDV	Destination Sequence Distance Vector
DSR	Dynamic Source Routing
DSR	Dynamic Source Routing
EDCA	Enhanced Distributed Channel Access
EDCAF	Enhanced Distributed Channel Access Function
EIFS	Extended IFS
ERP	Extended Rate PHY
ESS	Extended Service Set
FSR	Fisheye state routing
GPSR	Greedy Perimeter Stateless Routing
HC	Hybrid Coordinator
HCCA	HCF Controlled Channel Access
HCF	Hybrid Coordination Function
HEC	Header Error Check
HSR	Hierarchal State Routing
IFS	Inter Frame Spaces
LLC	Link Layer Control
MAC	Medium Access Control
MADV	Medium-Aware Distance Vector
MPDU	MAC Protocol Data Unit
MPR	Multipoint relays

MSDU	MAC Service Data Unit
MTM	Medium Time Metric
NAV	Network Allocation Vector
OLSR	Optimized Link State Routing
OSI	Open Systems Interconnections
PCF	Point Coordinated Function
PIFS	PCF IFS
PMD	PHY Media Dependent
PSDU	PLCP Service Data Unit
QoS	Quality of Service
RREP	Route Reply
RREQ	Route Request
RSI	Required Service Interval
RSSI	Received Signal Strength Indicator
RTS	Request To Send
SFD	Start Frame Delimiter
SI	Service Interval
SIFS	Short IFS
STA	(IEEE 802.11 conformant) Station
TDBRPF	Topology Dissemination Based on Reverse-Path Forwarding
TORA	Temporally-Ordered Routing Algorithm
TS	Traffic Stream
TSPEC	Traffic Specification
TTL	Time to Live
TXOP	Transmit Opportunity
WLAN	Wireless Local Area Network
WPAN	Wireless Personal Area Network
ZRP	Zone Routing Protocol

Chapter 1

Introduction

1.1 Motivation

The advent of wireless data communication has revolutionized our life styles. Through the evolutionary process of development, a myriad of network applications have been developed for using the network services where the last mile connectivity is through a wireless medium. The requirements of new applications and end users from the network technologies are becoming even more challenging with time. To cope up with such challenging demands, a constant and even thorough research is required for improving the existing protocols, devising new standards and technologies.

The research presented in this thesis is motivated by two, inter-linked issues:

1. The IEEE 802.11[1], family of standard specifications for medium access control (MAC) protocol and the physical layer (PHY) has a key role in the revolution of wireless data communications. Over a period of time since the first release of the IEEE 802.11 standard a series of enhancements have been introduced. The 802.11 standard and its later extensions are particularly focused at the specifications for MAC sub-layer and the PHY. However, the set of protocols which controls the end-to-end communication between a sender and a receiver in a network were initially designed for wired networks. With the replacement of wired PHY (and its associated Data-Link layer protocols) by wireless PHY and its associated MAC, the higher layer's protocols remained ignorant of the fact that now they are operating without a wire. The lack of knowledge about the working

mechanism of new protocols and more importantly about the different characteristics of the new PHY consequently became a cause of wrong assumptions at higher layers. It was reported widely through extensive research that in order to enhance the communication performance, the higher layers protocols should have a mutual information exchange with the new PHY and related protocols. While the layer-to-layer abstraction was the main goal of the layered architecture; in case of wireless networks it caused several performance bottlenecks. To overcome this issue, the idea of cross-layer design is proposed in several research papers. In a cross-layer design, parameters from protocols at various layers are exchanged with other protocols across the protocol suite. Cross-layer information exchange has been presented as an effective and inevitable solution in many cases [2-6].

The difference in the characteristics of the wireless (and an IEEE 802.11) PHY when compared with its wired predecessors comes from the fact that in case of a wireless PHY the communication is affected by presence of other stations in the vicinity causing higher contentions, presence of sources of interference causing transmission disruptions, the variation of transmission rates used, the distance between the transmitter and a receiver, and the mobility patterns of stations. All these factors affect the 'link quality' and thus the communication efficiency.

Appropriate mechanisms are required to detect changes in the link quality (and thus all the contributing factors) and then adjustments are made to the behavior of related protocols for performance efficiency under prevailing conditions. Transmission rate holds a great significance in determining the overall efficiency of a communication system. Generally, it is always desirable to use the highest transmission rate e.g. highest transmission rate can potentially yield: highest throughput, lower medium occupancy (thus lower contention and spectrum efficiency) and lower power consumption. An IEEE 802.11 conformant receiver can successfully receive transmitted frames if the signal to noise ratio which it experiences at the time of the reception is higher than a threshold value. Because of the underlying modulation schemes the sensitivity threshold increases for transmissions made at a higher rate. Therefore, with variation in the channel quality, higher rate transmissions are prone to erroneous reception as compared to lower rate transmissions. Ideally, for efficient communication, a station should reduce the transmission rate when its receiver is unable to receive and should increase the transmission rate whenever possible. This process of selection of appropriate transmission rate according to prevailing medium/communication status is called rate-adaptation.

Although, rate-adaptation protocol is of significant importance, the IEEE 802.11 standard does not provide standard specification for a rate-adaptation technique. The standard, however, specified mandatory rules for devising a rate adaptation technique. The lack of standard specification for a rate-adaptation scheme and its significance in determining the communication efficiency makes it an important issue and motivates for further research.

2. Mobile ad-hoc networks (MANETs) have received a tremendous attention from the research community. In contrast to wired networks, MANETs are characterized by dynamic topological changes due to station mobility; such a trait of MANET makes the task of establishing routes a very challenging process. Early efforts for MANET routing protocols were inspired by the methodologies used in routing protocols for wired networks. Such trends led to several MANET routing protocols, where route selection process essentially tried to select routes with minimum distance between the source and destination stations and so the 'number of hops' between the source and destination became a default routing metric.

   However wireless communication stations use a shared frequency band where access regulation to a particular frequency band is coordinated with the help of MAC protocols. The MAC protocol ensures that only one station accesses the medium at one point in time to avoid collisions. In situations of higher station density within the transmission range, stations are put on hold for longer duration. In case of the IEEE 802.11 standard compliant stations, they may have to freeze their backoff counters a number of times before they find an opportunity to transmit on the medium. Transmission errors and station density accounts for higher medium occupancy by neighboring stations and as a result every station experiences higher contention delay for accessing the medium. Likewise, stations using lower transmission rate can penalize neighboring stations by causing higher medium contention. The end-to-end contention delay can potentially increase with increase in the number of stations along the route. Moreover, a routing protocol which selects a route by including links with higher transmission rates would generally imply that the route has a higher number of intermediate stations between a source and destination. Higher number of intermediate stations means relatively higher contention delay along the route. In addition, in a multi-rate, multi-hop network, the end-to-end throughput is capped by a station using a relatively lower transmission rate along the route. If lower transmission rate links follow higher transmission rate links, frames would pileup in the transmission queues at a station transmitting at a lower rate and as a result there would be higher queuing delay [7]. The same is true when packets, on its way through the multi-hop route, land in a highly congested area. Avoiding, congested (having higher medium access contention delays) links and constructing a route with relatively less contention and queuing delays necessitates incorporation of MAC and PHY specific information in the routing protocol.

Link quality parameters (e.g. transmission rate, frame error rate, signal to noise ratio) are rooted at the physical layer of the OSI architecture. Likewise, monitoring the medium contention levels, frame prioritization and queuing delays are dealt at the medium access control (MAC) sub-layer. The need of inclusion of PHY and MAC level information in the routing protocol as mentioned above makes such a design a classical case of cross-layer architecture. For an efficient design of a routing protocol for MANETs, such a cross layer design is inevitable and requires extensive research for formulating a design strategy so that MAC and PHY specific information is incorporated in the decisions of routing protocols.

1.2 Aims of research

The aims of the research presented in this thesis are mainly two folds:

1. To design an efficient rate-adaptation scheme for multi-rate capable wireless PHYs. The rate-adaptation scheme should be able to detect accurate channel state information and appropriately select a transmission rate which is the most suitable for communication under the prevailing channel quality.

   The rate-adaptation scheme should be highly responsive and should quickly react to variations in channel quality. Retransmissions and other forms of communication overheads should be minimized and if possible avoided altogether. The proposed scheme should be able to detect true causes of channel quality variations; therefore, it is aimed to have an effective mechanism for determining causes of frame losses.

   Most importantly, the rate-adaptation scheme should require no modification to the standard frame formats so that compatibility miss-matches are avoided.

2. Secondly, the research aims at a design of a MANET routing protocol where the route selection process takes into account the communication constraints and capabilities of every station along the route and selects a route which is most appropriate according to the Quality of Service (QoS) demands of outbound traffic.

   The route selection metric should consider the medium access delays and the transmission rate used by stations in a network. In order to successfully exploit the QoS facilities introduced by the IEEE 802.11e, the targeted routing protocol should take into consideration the variation of unfairness in queuing delays among the four access-categories (AC) queues.

   It is very important to notice that the above mentioned, MAC specific parameters are highly variable; therefore, an efficient mechanism of adjusting the behavior of the routing protocol according to the variations in the underlying MAC-parameters should be included in the routing protocol.

1.3 Contribution to knowledge

This thesis contributes to knowledge by giving a detailed insight into various rate-adaptation techniques, analyses various aspects of such protocols and focuses on their implications on the performance of communication. The findings are then used to design two of rate-adaptation schemes, each belonging to a different class of rate-adaptation approaches, namely: the sender-side statistics based rate-adaptation and receiver-side SNR based rate adaptation.

As discussed earlier, MAC level enhancements have a significant impact on the performance of higher layers. In order to demonstrate the significance of its impact and the advantages of using cross-layer information, our next contribution primarily focused on the Network Layer; wherein we proposed a MANET routing protocol which is based on a novel idea and incorporates MAC (and in turn PHY) specific parameters in the routing metric.

The key contributions are summarized as follows:

1. A frame-failure-statistics based rate adaptation solution which uses an *on-demand incremental* strategy for selecting a rate-selection threshold. This solution is based on a cross-layer communication framework, where the rate-adaptation module involves information to/from application layer along with relevant information from the medium access control (MAC) sub-layer while making a decision.
    a. The on-demand incremental strategy avoids the chances of retransmissions.
    b. The rate-control function works in close coordination with the application requirements.
    c. The underlying MAC sub-layer’s timing constraints and medium congestion is taken into account while performing rate-adaptation.
    d. A feedback mechanism is used to communicate the underlying layers’ status and capability information to higher layers.
    e. A novel mechanism of frame-loss differentiation is introduced to distinguish between frames lost because of erroneous transmissions (lost frames) from frames which are not received because of simultaneous transmission by other stations.

2. An SNR-based rate-adaptation scheme, called MutFed, relying on mutual feedback between a transmitter and receiver pair.
    a. MutFed uses a novel mechanism of conveying feedback between communication peers, without the need for changing the Standard frame formats.
    b. The proposed rate-adaptation scheme is highly responsive to variations in communication-channel's quality.
    c. MutFed uses an extremely efficient and highly responsive frame-loss differentiation mechanism for distinguishing *corrupted frames* from *collided frames.*
3. A Medium Aware Distance Vector (MADV) routing protocol. MADV formulates a routing metric which combines variations in the transmission rates, contentions levels and queuing delay at every link. A significant addition of MADV is the incorporation of the queue prioritization information which is introduced by the IEEE 802.11e MAC specification. The IEEE 802.11e introduces four different MAC level queues for each of the four access categories. The standard specifies maintenance of service priority within the queues; which implies that frames from a higher priority queue would be serviced more frequently than frames belonging to lower priority queues. Such an enhancement at the MAC sub-layer introduces uneven queuing delays. Conventional/existing routing protocols are unaware of such queue prioritization and as a result these factors are not considered which result in severe performance deterioration for frames belonging to lower priority queues. It is important to consider hop-by-hop (or a link) MAC level information while constructing an end-to-end routing strategy for a multi hop wireless network. In MADV the routing table structure is modified from the existing routing protocol's approaches by introducing per destination-per AC entry; which essentially means that a source station can have different route entries based on the ACs for the same destination. The advantage of ACs-specific route entries is that MAC-level queue prioritization can be effectively taken into account for the whole route, and the route selection is performed based on the lowest metric value per AC from the available routes to a destination.

The research which led to the contribution of MADV can be summarized to include the following contributions:

a) Investigation of the significance of variation of the transmission rate, the use of a rate-adaptation scheme at the MAC layer and its effects on the performance of routing protocols.
b) The significance of medium occupancy, PHY and MAC level factor contributing to this phenomenon and the effects of medium contention on the performance of route selection.
c) An in depth analysis of the IEEE 802.11e inter-queue access priority, the resultant queue delay variations and their effects on routing strategy. These three factors lead to the formulation of the MADV routing metric.
d) Formulation of cross-layer information exchange frame work for communication MAC and PHY specific parameters for inclusion in the routing protocol route metric.
e) A novel route selection mechanism for MANETs, called MADV, which incorporates the underlying MAC and PHY specific parameters and the overall route selection, is based on the actual state of the medium.

1.4 Research methodology

The research methodology used for conducting the research presented in this thesis, is summarized as follows:

1. An in-depth analysis of the capabilities of the IEEE 802.11 standard conformant devices is presented to enable the formulation of rate-adaptation techniques and inclusion of MAC specific parameters in the MADV. For compliance with the standard, a review of various extensions of the standard was done in the initial phase of the research.
2. A comprehensive analysis of various published articles focusing on the rate-adaptation techniques. The analysis gave an insight into various issues which needed to be included in the proposed techniques.

3. A review of various MANET routing protocols, specifically those where the research focus was on the inclusion of MAC specific parameters.
4. Design of proposed rate-adaptation schemes and the MADV routing protocol.
5. Development of a simulation model for the proposed rate-adaptation schemes and the routing protocols. The simulation models were developed in OPNET [8] network simulator. The model development phase involved working at the 'Process modeling' level in the simulator; where a new functionality or a protocol is implemented with the help of a Proto-C programming language.
6. Development of simulation models for several existing rate-adaptation protocols. The simulation model for modeling the IEEE 802.11 standard MAC does not include a rate-adaptation scheme. Therefore, before commencing with development of the proposed models, we developed simulation models for several existing rate-adaptation schemes.
7. Validation of the developed models, tests of individual functions of the protocols in the simulation environment.
8. Performance tests and comparison of the proposed solutions with existing techniques.
9. Analysis of the results from the comparisons.

1.5 Thesis structure

This thesis consists of six chapters. Each chapter is structured independently. Conceptually, the chapters are inter-dependent and the reader should follow the right order in order to better understand the contributions presented in the thesis.

After the introductory chapter-1, the chapter-2 gives a brief overview of the IEEE 802.11 standard specification specifically focusing on the issues which will be referred to in the later chapters. A summary of the various enhancements which were later introduced to the initial release of the standard is given, followed by a brief overview of fundamental concepts and terminologies used in the standard. A great deal of emphasis is given to working mechanisms of DCF, PCF and HCF coordination functions along with the PHY dependent characteristics. The MAC level traffic prioritization introduced by the IEEE

802.11e is explained in detail along with the two modes of the HCF i.e. the EDCA and HCCA. The design constraints on devising rate-adaptation strategies are elaborated and given in detail in chapter-2 so as to consolidate the concepts which would be used in later chapters in connection to the IEEE 802.11 multi-rate operation capability. Conceptually, chapter-2 is a pre-requisite for the later chapters.

Chapter-3 gives a detailed insight into various rate-adaptation techniques, analyses various aspects of such protocols and focuses on their implications on the performance of communication. This chapter presents two rate-adaptation schemes. Various features associated generally with all rate-adaptation schemes, and specifically with the proposed rate-adaptation mechanisms are discussed in detail in this chapter. This chapter also gives a detailed insight into various MAC specific issues and forms the basis for chapter-5.

The designs presented in the chapter-3 are elaborated further with the help of performance tests and analysis in the following chapter-4.

Chapter-5 discusses in detail various classes of MANET routing protocols. It then gives an overview of the MAC and PHY constraints in communication and various factors which arise from these constraints and the effects of these factors on routing protocols, followed by a comprehensive analysis of various research efforts focusing on enhancements to routing protocols in connection with incorporation of medium awareness. Chapter-5 then gives detail of the MADV routing protocol, and explains the comparative analysis of the MADV routing protocol in various simulation scenarios.

Finally, Chapter-6 concludes the research findings of the thesis and presents the future work to be carried out in connection with the research presented in this thesis.

1.6 References

[1] "IEEE Standard for Information technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," IEEE Std 802.11-2007 (Revision of IEEE Std 802.11-1999), vol., no., pp.C1-1184, June 12 2007.

[2] S. Shakkottai, T. Rappaport and P. Karlsson, "Cross-layer design for wireless networks," Communications Magazine, IEEE, vol. 41, pp. 74-80, 2003.

[3] V. Srivastava and M. Motani, "Cross-layer design: a survey and the road ahead," Communications Magazine, IEEE, vol. 43, pp. 112-119, 2005.

[4] M. van Der Schaar, N. Sai Shankar, "Cross-layer wireless multimedia transmission: challenges, principles, and new paradigms," Wireless Communications, IEEE , vol.12, no.4, pp. 50-58, Aug. 2005

[5] N. Nang, R. Sankar and Jungsik Lee, "Improving ad hoc network performance using cross-layer information processing," pp. 2764-2768, 2005.

[6] V. Kawadia and P. Kumar, "A cautionary perspective on cross-layer design," Wireless Communications, IEEE [See also IEEE Personal Communications], vol. 12, pp. 3-11, 2005.

[7] X. Chen, H. M. Jones and A. D. S. Jayalath, "Congestion-Aware Routing Protocol for Mobile Ad Hoc Networks", 66th IEEE Conference on Vehicular Technology. VTC-2007. pp. 21-25, 2007.

[8] OPNET Modeler, OPNET Technologies Incorporation. www.opnet.com

Chapter 2

Operational Characteristics of the IEEE 802.11 MAC and PHY

2.1 Introduction

The research presented in this thesis deals with the design of rate-adaptation schemes for the IEEE 802.11 standard, and the effect of enhancements at the MAC sub-layer (as defined in the IEEE 802.11 standard specifications specifically and MACs defined in other standards generally) on the performance of routing protocols for ad-hoc mobile wireless networks. Before proceeding into the research presented in the following chapters, it is logical to present and highlight the core concepts and fundamental principles that are used in the design of the IEEE 802.11 standard specification for MAC and PHY.

The main goal of a MAC protocol is to regulate access of wireless devices to a shared wireless medium. In the process of doing so, several timing constraints are imposed by the MAC protocol in order to better regulate the shared resource and avoid collisions. Likewise, there are several factors, e.g. interference from neighbouring stations, presence of hidden stations, distance between the sender and the receiver, and the number of stations contending for medium access; all of these factors are taken in account by the MAC protocol and enforces its rules accordingly. As a result of which the communication performance is highly variable.

Over a period of time since the first release of the IEEE 802.11 standard a series of enhancements some addressed at the MAC and other at the PHY have been introduced. These enhancements have further complicated the design and working of the MAC protocol and enhancements techniques that are based at the MAC or at the higher layers need to take such stringent constraints in consideration.

Likewise, the IEEE 802.11 family of standard (and several extensions) comprises of several PHY. Each of the PHY is different in its operational parameters and characteristics. There are several MAC level parameters that are dependent on the PHY's characteristics. Therefore, it is also necessary to have a detailed understanding of the various characteristics of the several PHYs.

This chapter gives a brief overview of the IEEE 802.11 standard specification specifically focusing on the issues which will be referred to in the later chapters. A summary of the various enhancements which were later introduced to the initial release of the standard is given, followed by a brief overview of fundamental concepts and terminologies used in the standard. A great deal of emphasis is given to working mechanisms of DCF, PCF and HCF coordination functions along with the PHY dependent characteristics. The MAC level traffic prioritization introduced by the IEEE 802.11e is explained in detail along with the two modes of the HCF i.e. the EDCA and HCCA. As mentioned later in the following chapter the IEEE 802.11 standard does not specify a rate-adaptation scheme, however it does specify some fundamental principles which are mandatory for designing a rate adaptation scheme; such design constraints are elaborated and given in detail in this chapter so as to consolidate the concepts which would be used in later chapters in connection to the IEEE 802.11 multi-rate operation capability.

2.2 Introduction to the IEEE 802.11 Standard- Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specification

This section provides a general description of the IEEE 802.11 WLAN standard specification for MAC and PHY. This is a fairly long section, covering specific aspects of the IEEE 802.11 MAC specification which will be used as basis for discussion later in chapter-3, 4 and 5. This chapter provides detail of the enhancements to the original 802.11 standard along with the fundamental concepts which are involved in the very successful and revolutionary standard for WLANs.

In 1997, the Institute of Electrical & Electronics Engineers (IEEE) approved the standard specification for wireless local area networks (WLANs) under the identification number of 802.11 [1]. Since the initial release of the IEEE 802.11 standard, a number of extensions have been approved during the last eleven years. These extensions enhanced the PHY and MAC layer capabilities of the original 802.11 standard. A summary of various extensions of the IEEE 802.11 standard on a timeline is given in table 2-1.

Table 2-1: The IEEE 802.11 standard and series of extensions since the first release.

Extension	Timeline	Status	Description	Detail
802.11 [1]	1997	approved	Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specification	The medium access control (MAC) and physical (PHY) characteristics for wireless local area networks WLANs are specified in this standard. This standard contains three physical layer units: two radio units, both operating in the 2400-2500 MHz band, and one baseband infrared unit.
802.11a [2]	1999	approved	High-speed Physical Layer in the 5 GHZ Band	Provisions to support the new high-rate PHY for operation in the 5 GHz band.
802.11b [3]	1999	approved	Higher-Speed Physical Layer Extension in the 2.4 GHz Band	Provisions to support the higher rate PHY for operation in the 2.4 GHz band.
802.11d [4]	2001	approved	Specification for operation in additional regulatory domains	Specifications for conformant operation beyond the original six regulatory domains of that standard. These extensions provide a mechanism for an IEEE 802.11 access point to deliver the required radio transmitter parameters to an 802.11 mobile station, which allows that station to configure its radio to operate within the applicable regulations of a geographic or political subdivision. This amendment enables the 802.11 mobile stations to roam between regulatory domains.

802.11g [5]	2003	approved	Further Higher Data Rate Extension in the 2.4 GHz Band	Provision of support for further higher data rates for operation in the 2.4 GHz band.
802.11h [6]	2003	approved	Spectrum and Transmit Power Management Extensions in the 5 GHz band in Europe	IEEE Std 802.11h-2003 provides mechanisms for dynamic frequency selection (DFS) and transmit power control (TPC) that may be used to satisfy regulatory requirements for operation in the 5 GHz band in Europe.
802.11i [7]	2004	approved	Medium Access Control(MAC) Security Enhancements.	Security mechanisms for IEEE 802.11 are defined in this amendment. The Temporal Key Integration Protocol (TKIP) and CTR with CBC-MAC Protocol (CCMP) mechanisms defined here provide more robust data protection mechanisms than WEP affords.
802.11j [8]	2004	approved	4.9 GHz–5 GHz Operation in Japan	This amendment specifies the extensions to IEEE Std 802.11 for WLANs providing mechanisms for operation in the 4.9 GHz and 5 GHz bands in Japan.
802.11e [5]	2005	approved	Medium Access Control (MAC) Quality of Service Enhancements	This amendment defines the MAC procedures to support WLAN applications with quality of service (QoS) requirements. The procedures include the transport of voice, audio, and video over IEEE 802.11 WLANs.
802.11k [9]	2008	active	Radio Resource Measurement of Wireless LANs	The proposed Radio Resource Measurement approach is to add measurements that extend the capability, reliability, and maintainability of WLANs through measurements and provide that information to upper layers in the communications stack. The Radio Resource Measurement scope is to define Radio Measurements and to provide mechanisms to higher layers for radio and network measurements. This amendment provides these mechanisms using request/response queries and an Object Identifier (OID) interface to upper layers in the Management Information Base (MIB).
802.11r [10]	2008	active	Fast Basic Service Set Transition (FT)	This amendment describes mechanisms that minimize the amount of time data connectivity is lost between the station (STA) and the distribution system (DS) during a basic service set BSS transition. The STA determines when to transition and to which access point (AP) to transition based on a number of factors, some of which may be out of the scope of this standard.
802.11n [11]	2008	active	Enhancements for Higher Throughput	This amendment defines modifications to both the 802.11 PHY and the 802.11 MAC so that modes of operation can be enabled that are capable of much higher throughputs, with a maximum throughput of at least 100Mb/s, as measured at the MAC data service access point (SAP).
802.11u [12]	2008	Active	Interworking with external networks	This amendment specifies enhancements to the 802.11 MAC to support WLAN interworking with external networks.

2.3 General description of the architecture and services.

The IEEE 802.11 introduces several architectural components which interact to form a WLAN. A conceptual diagram highlighting various components of an IEEE 802.11 based WLAN is shown in Figure 2-1.

2.3.1 Portable and Mobile Stations (STA):

A device having IEEE 802.11 conformant MAC and PHY is called an STA. The standard supports *portable* and *mobile* STAs. A *portable* STA is used only at fixed position and cannot operate while it is on the move. On the other hand a *mobile* STA is capable to communicate while moving.

2.3.2 Basic service set (BSS):

A set of STAs which are all governed by a single (the same) coordination function is called a basic service set. A *Coordination function* is a logical set of rules which determine and regulate access to shared wireless medium. A BSS can be controlled by a Hybrid Coordination Function (HCF), or it can have an HCF along with a point coordination function (PCF) and will have an obligatory distributed coordinated function (DCF).

2.3.3 Independent BSS (IBSS):

A BSS which is not connected to a *distribution system* (DS) is called an independent BSS. An IBSS can be formed on the fly, with the minimum number of two STAs and for this reason such an IBSS is also termed as an *ad hoc* network.

2.3.4 Distribution system (DS):

A distribution system is one of the several architectural components defined by the IEEE 802.11 standard for connecting multiple BSSs. A DS provides logical services for handling address-to-destination mapping and seamless integration of multiple BSSs. An Access Point (AP) provides access to a DS for all associated STAs.

2.3.5 Extended Service Set (ESS):

An ESS is a group of BSSs which are connected through a DS. The DS is only used as a means of connecting partially/completely overlapping or completely disjoint BSSs and is not included as an architectural component of an ESS. The standard does not define or limit the physical locations of BSSs within an ESS. All the BSSs within an ESSs appear like a single IBSS to the Link Layer Control (LLC) layer and mobile STAs can move from one BSS to another transparently to the LLC.

2.3.6 QoS BSS:

A QoS BSS supports LAN applications with QoS requirements through MAC sub-layer enhancements (introduced through IEEE 802.11e). The standard has two provisions for supporting applications with QoS requirements.

An *Enhanced Distributed Channel Access (EDCA)* achieves QoS support by prioritizing frames. Frames are categorized and every category has its own priority parameters. A frame's priority is established by varying:

- The amount of time that a STA senses the channel to be idle before transmitting a frame.
- The length of contention window to be used for backoff.
- The duration of time that a STA will use for transmission after acquiring access to medium.

Another mechanism for supporting MAC level QoS is called *Hybrid Coordination Function (HCF) Controlled Channel Access (HCCA).* Under the HCCA scheme, a *Hybrid Coordinator (HC)* which is collocated at the *Access Point (AP),* allows allocation of *Transmission Opportunities (TXOP)* to QoS STAs. QoS STAs request for reservation of TXOPs with the HC for transmission of frames from itself to the AP and for frames from the AP to itself. The HC allocates or rejects the requests for TXOP allocation based on the outcome of an admission control scheme. When a TXOP is allocated to a QoS STA, the HC polls the QoS STA according to the parameters conveyed by the QoS STA when requesting for TXOP.

The QoS enhancements are available through the HC which is collocated at the AP, therefore, in order to avail the full QoS enhancements there should be an AP supporting the HC functionality in a BSS. However, a subset of the QoS enhancements can be used by QoS STAs without the presence of an AP (having an HC), forming a QoS IBSS. The QoS IBSS uses HCF where the TXOPs are obtained through the EDCA mechanisms. All the parameters which control the EDCA operations, which are sent by an HC when a QoS BSS is operating, are fixed in this case. Furthermore, a QoS IBSS does not support polled TXOP and Traffic Specification (TSPEC) cannot be established due to the absence of an HC.

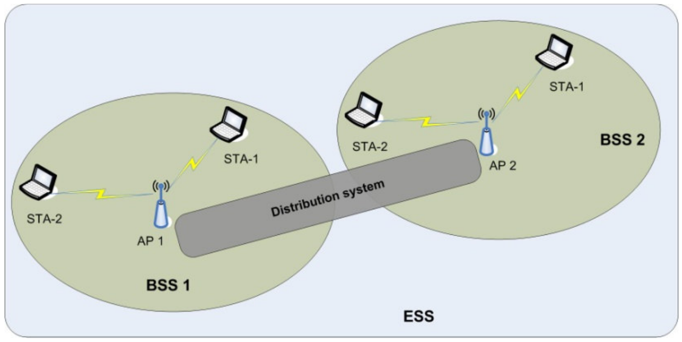

Figure 2-1: Components of the IEEE 802.11 architecture

2.3.7 Mobility support:

There can be various forms of STA mobility depending on the logical and/or physical position of a STA within the WLAN.

1. Mobility within a BSS: A mobile STA moves within the coverage area associated with its BSS.

2. Mobility within an ESS: A mobile STA moves from one BSS to another of the same ESS. The standard supports this mobility and transition is performed transparently to the LLC.

3. Inter ESS mobility: A mobile STA moves from one ESS to another ESS. In this case, the transition is no longer transparent to the LLC and as a result may cause disruption of connections at higher layers.

2.4 IEEE 802.11 specification for MAC sub-layer

2.4.1 Coordination Functions of the IEEE 802.11 MAC protocol:

The standard specifies a mandatory coordination function for regulating channel access among STAs which operate within a BSS. The mandatory coordination function of the IEEE 802.11 MAC is called Distributed Coordination Function (DCF). A *Hybrid Coordination Function (HCF),* operates by using the services of the DCF. HCF has three components i.e.

1. The HCF contention access mechanism called *Enhanced Distributed Channel Access (EDCA),* for prioritized channel access.

2. The HCF controlled channel access (HCCA) mechanism for parameterized QoS services through the MAC sub-layer.

3. A Point Coordination Function (PCF) for contention free channel access.

The HCF enhancement is only present within STAs in a QoS BSS. PCF is optional for QoS and non-QoS BSS. Figure 2-2 shows the MAC architecture.

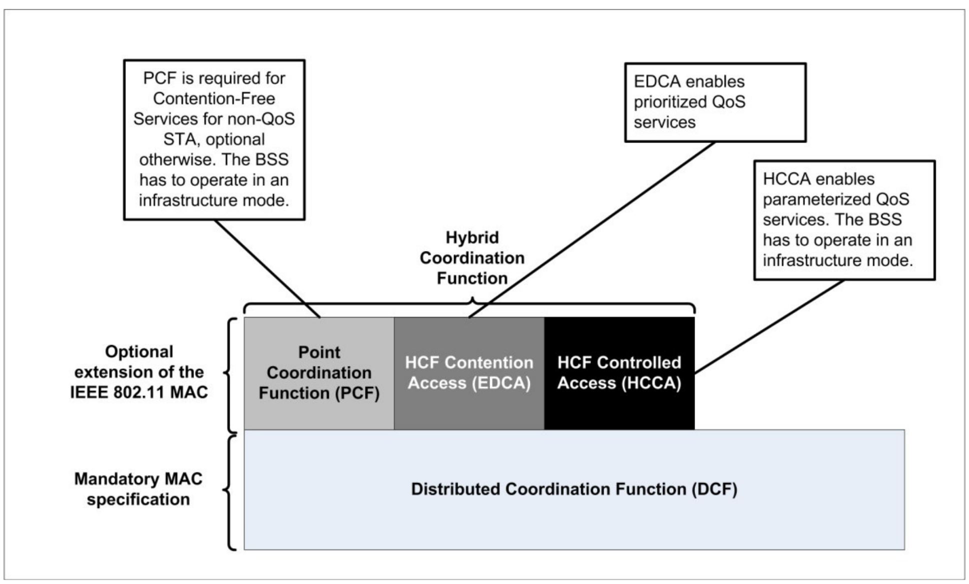

Figure 2-2: IEEE 802.11 MAC architecture, optional and mandatory components.

2.4.2 Carrier Sensing (CS)

The IEEE 802.11 MAC protocol regulates access to a shared wireless medium among STAs using the *carrier sense multiple access with collision avoidance (CSMA/CA)* mechanism. Carrier sensing (CS) is performed through two ways:

- PHY-CS: The medium status is determined through the information passed in the PHY Clear Channel Assessment Indication (PHY-CCA.Indication) parameter. The PHY provides information about the existing medium status to the MAC sub-layer through this parameter.

- Virtual-CS: The virtual CS is performed by distributing medium occupancy information within a BSS. Medium occupancy information can be propagated across a BSS either through the use of Request-to-Send/Clear-to-Send (RTS/CTS) procedure or by using information in the 'Duration ID' field in individually addressed frames. The 'Duration ID' field specifies the amount of time when the medium is occupied by a STA. This duration is either till the end of current frame exchange which is marked by the receipt of acknowledgment (ACK) frame, or in case of fragmented MSDU, the duration ID indicates time when an ACK frame for the next frame is received. Every STA uses the virtual-CS to maintain a Network Allocation Vector (NAV). The NAV is a counter, whose value is set through the virtual-CS and is decremented with the passage of time. The NAV indicates the ongoing (near future) usage of the medium; a zero value of the NAV implies that the medium is *idle* according to the virtual-CS.

2.4.3 Inter Frame Spaces (IFS):

To avoid collision (simultaneous access by STAs) the CS procedure is used for different intervals at various instants of time during frame exchange. The CS intervals are commonly called *Inter Frame Spaces (IFS)*. The standard specifies five IFSs:

1. *Short IFS (SIFS):*

   The SIFS is used before transmitting an ACK frame, or a CTS frame after receiving an RTS, or before transmission of subsequent MPDUs during fragment burst mode or before responding to a PCF poll.

2. *PCF IFS (PIFS):*

   The PIFS is used by STAs operating under PCF for gaining medium access at the start of contention free period (CFP). Likewise it is used by STA before transmitting a Switch Channel Announcement frame.

3. *DCF IFS (DIFS):*

   The DIFS is used in DCF, wherein a STA is allowed to transmit if the CS indicates 'idle' status for DIFS-interval and the backoff counters (see the next section), becomes zero.

4. *Arbitration IFS (AIFS):*

   The AIFS is used in for channel access in a QoS BSS. To meet QoS requirements, the MAC prioritizes access of some Access Categories (ACs) by using variable AIFS intervals. An AIFS duration is defined as:

$$AIFS[AC]=AIFSN[AC] \times aSlotTime^{1} + SIFS \tag{2.1}$$

   AIFS Number (AIFSN) is different for each AC, for example the default values used by the non-AP STAs for AIFSN are 7, 3, 2, 2 for AC_BK[2], AC_BE[3], AC_VI[4] and AC_VO[5] respectively. Therefore, the corresponding AIFS would be of different interval for each AC which implies that a STA's wait time is variable according to the type of AC. See Figure 2-3 for detail of AIFS timing relationship.

5. *Extended IFS (EIFS)*

   There are times when transmission between two STAs is disrupted by neighbouring STA in such a way so that these two STAs are unable to decode each other's transmission. When erroneous frames are received at STAs they should wait for a longer time before proceeding with next frame transmission. This is used as a precaution to avoid any further collisions, and providing the interfering STAs enough time to complete their ongoing transmission. In such cases, STAs wait for EIFS duration when the previous received frame is received in error.

---

[1] aSlotTime is a time unit used by the IEEE standard as a reference for measuring time. The value of aSlotTime is PHY dependent.

[2] Traffic categorized as of the type *BACKGROUND.*

[3] Traffic categorized as of the type *BEST EFFORT.*

[4] Traffic categorized as of the type *VIDEO.*

[5] Traffic categorized as of the type *VOICE.*

The relationship between various inter-frame spaces as given in the standard can be explained with the help of the following equations:

$$SIFS = aRxRFDelay + aRxPLCPDelay + aMACProcessingDelay + aRxTxTurnaroundTime \quad (2.2)$$

$$SlotTime = aCCATime + aRxTxTurnaroundTime + aAirPropagationTime + aMACProcessingDelay \quad (2.3)$$

$$PIFS = SIFS + SlotTime \quad (2.4)$$

$$DIFS = SIFS + 2 \times SlotTime \quad (2.5)$$

$$EIFS=SIFS + DIFS + ACKTxTime \quad (2.6)$$

The parameters which are used to define various inter-frames spaces in the above given equations represent the following time values; it is important to notice that some of the parameters are entirely PHY dependent while others are dependent on the MAC which makes some of the inter-frames spaces dependent both on the design of PHY and MAC.

- The '*aRxRFDelay*' is the nominal time in microseconds which represents the duration between the end of a symbol at the air interface to the moment when the PMD indicates (interrupts or signals) the PLCP about the arrival of data. This essentially represents the time between when the reception of data is completed at the receiver and till the time the PLCP is notified of such reception.
- The '*aRxPLCPDelay*' represents that time which is used by the PLCP to deliver whatever it has received from the PMD to the MAC. This value is also indicated in microseconds.
- The '*aMACProcessingDelay*' shows the time in microseconds which is used by a MAC to change the mode either for transmission or CCA.
- The '*aRxTxTurnaroundTime*' represents the maximum amount of time required by the PHY to change from reception to transmission mode.
- The '*aCCATime*' indicates the minimum amount of time which is available for the CCA mechanism to sense the medium.
- And finally the '*aAirPropagationTime*' represents twice the propagation time which is required for a signal to travel to the most distant allowable station. For most of the PHYs the value of this parameter is either 1µs or less.

- ACKTxTime is the time expressed in milliseconds required to transmit an ACK frame, including preamble, Physical Layer Convergence Protocol (PLCP) header and any additional PHY dependent information, at the lowest PHY mandatory rate.

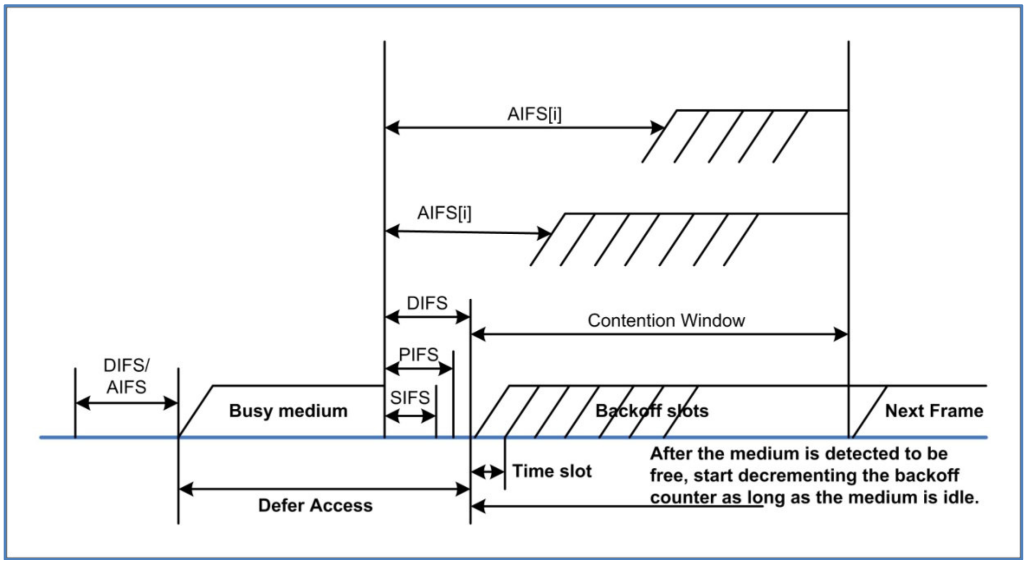

Figure 2-3: Relationship of various inter-frame spaces and their operational significance in the IEEE 802.11 MAC.

2.4.4 Random backoff time:

To minimize the chances of collision, the standard uses a random backoff interval for every station. When the CS mechanism reports an *idle* status for DIFS or EIFS (after a previous erroneous frame) interval a STA waits for random number of time slots. The backoff is calculated as:

$$\textit{Backoff time} = \textit{Random}\ ()\ x\ \textit{aSlotTime} \tag{2.7}$$

The *Random()* function uses a uniform distribution over an interval [0, CW], and generates a random integer value. The *Contention Window (CW)* has a value between [CW_{min}, CW_{max}], with the initial value set to CW_{min}. The CW is incremented to a next higher value for a retransmission. The value of CW is initially set to a PHY dependent CW_{min} and the increments are sequentially to the power of 2 minus 1.

2.4.5 DCF access procedure:

A STA while operating under the DCF procedure monitors the medium status using the CS mechanisms; given that an MPDU is in the transmission queue, the STA waits until the medium is idle for DIFS interval. The STA then checks it backoff counter and if the backoff counter has a non-zero value, the STA waits for that many number of timeslots. If the medium status changes from *idle* to *busy* while the STA performs backoff, it freezes the backoff procedure and waits till the medium becomes *idle* again. If the backoff counter has a zero value, the immediate access to the medium is allowed after DIFS interval. The DCF operational procedure is explained in Figure 2-4.

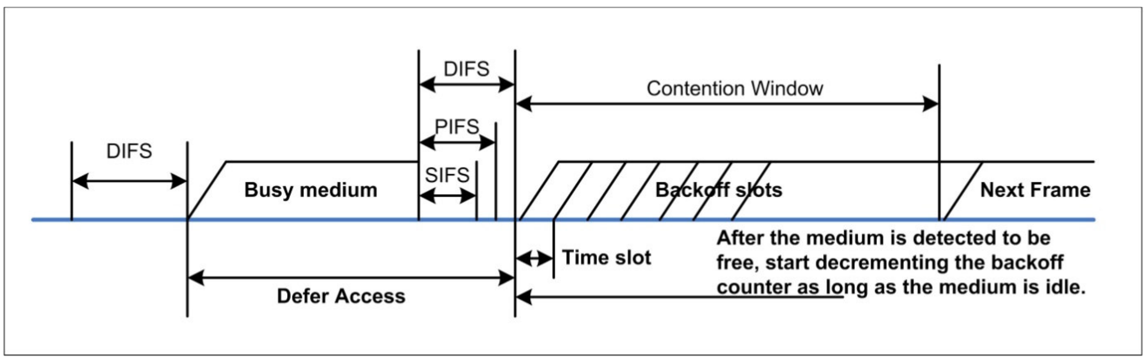

Figure 2-4: Distributed Coordination Function's basic access procedure.

2.4.6 Fragment bursting:

Transmission of maximum size MSDUs minimizes the transmission overhead and maintains higher throughput, however, in situation when the frame loss probability is high, large size MSDUs would imply higher loss. Therefore, the IEEE 802.11 standard allows fragmentation of MSDUs. A fragmentation threshold is defined to limit the size of an MPDU; therefore, when the size of an MSDU is larger than the fragmentation threshold, the MSDU is fragmented and every MPDU is equal to the fragmentation threshold value.

A STA can send multiple fragments of an MSDU after acquiring access to the medium, this is called fragment burst. In this mode, the sender waits for ACK for every fragment, and then transmits the next fragment after an SIFS interval. The fragment burst mode is explained with the help of Figure 2-5.

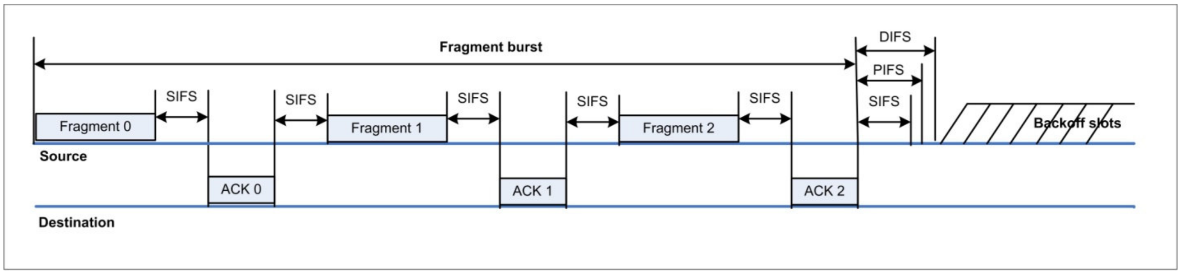

Figure 2-5: Transmission of multiple fragments of an MSDU separated by SIFS. This mode of operation is also called fragment burst.

2.4.7 PCF Access Procedure:

The PCF mechanism enables the point-coordinator (PC) to have priority access to the medium by waiting for relatively shorter duration (the PIFS duration) as compared to the DIFS. Under the PCF procedure, Contention Free Periods (CFPs) are generated by the PC, followed by contention periods (CPs). In order to be included in the CFP polling list of a PC (where the PC polls every STA for transmission of frames in a CF manner), every STA indicates its CF-pollability at the time of association/re-association. A PC can use the CFP for delivery of frames to and from itself. In the later case (when it uses CFP only to deliver frames to STAs) the PC does not need to maintain a polling list. At the beginning of a CFP, the PC senses the medium and if found idle sends a beacon frame. After sending a beacon frame the PC waits for SIFS and then send either a CF-poll, Data+CF-poll, management or if it doesn’t have any polling list entries and no data frames for STAs, it can send CF-end frame. STAs respond after an SIFS interval to a frame received from the PC during the CFP. The PC informs about the maximum CFP duration in every beacon frame. At the beginning of a CFP, all STAs set their NAVs to the MaxCFPDuration. When the PC ends CFP before the MaxCFPDuration (by sending a CF-End) these stations then reset their NAVs. The PCF access procedure is explained in Figure 2-6 as given in [5].

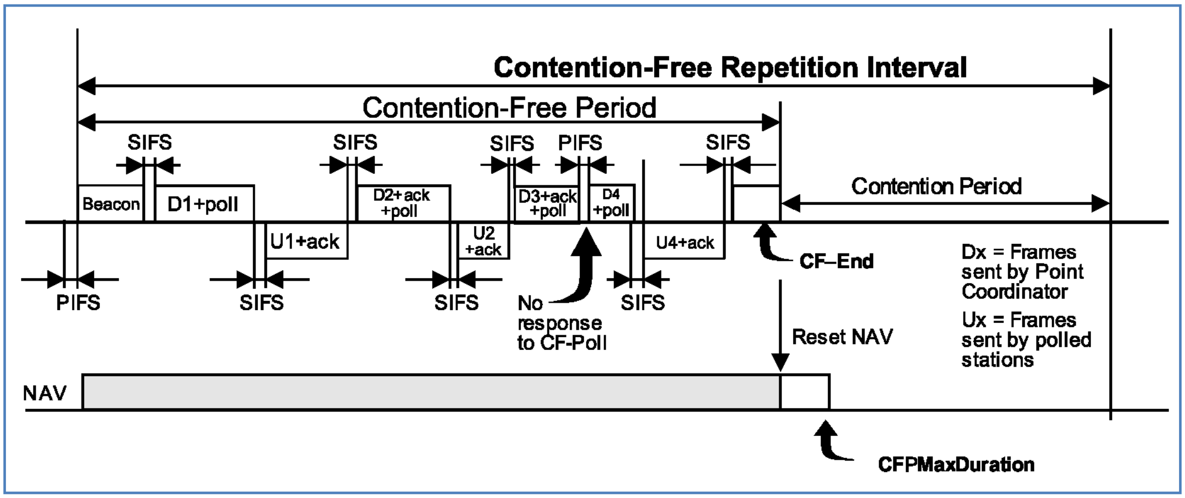

Figure 2-6: Point Coordination Function access procedure.

2.4.8 HCF

The HCF defines a contention based channel access mechanism called EDCA, and a controlled channel access called HCCA. HCF introduces the concept of *transmission opportunity (TXOP);* a TXOP is a unit of permission of accessing the medium. A TXOP can be obtained either through the EDCA or the HCCA procedure.

2.4.8.1 EDCA

The EDCA mechanism supports differentiated and prioritized access at the MAC sub-layer by using eight User Priorities (UPs) specification of the MSDUs. Each MSDU is tagged with its associated user priority when it enters the MAC sub-layer. The tagging of frames before they enter the MAC sub-layer with different priorities is implementation specific. The IEEE 802.1D [13] specifies various traffic classes and associated User Priorities. The UP values are guidance to the underlying MAC protocols to provide differentiated service to the traffic classes based on the UP values. The UP values are not changed at the MAC layer, unless there is a need to map them to MAC specific values. Therefore, the UP values hold an end-to-end significance [14]. The UP values are associated to various traffic classes; a higher value of the UP indicates a higher priority. The following explanation of various traffic classes and their associated Ups (and their corresponding UP numbers) is given with further elaboration in [14, 15].

- **Network control (7)**: Associated with traffic classes where the traffic is for network control and operation. This is a higher priority traffic class, which implies that traffic belonging to this class is time critical and its delivery is of higher priority.

- **Voice (6):** This traffic class is associated with traffic generated by interactive voice application which is delay sensitive. The delay requirement for this traffic class is 10-ms delay. As mentioned below, the UP for Voice is directly mapped to an access category (AC VO) Voice at the MAC sub-layer of the IEEE 802.11e.

- **Video (5):** This traffic class is associated with traffic generated by interactive video application (video conferencing) which is delay sensitive. The delay requirement for this traffic class is 100-ms delay. As mentioned below, the UP for Video is also directly mapped to an access category (AC VI) Video at the MAC sub-layer of the IEEE 802.11e.

- **Controlled load (4):** Non-time-critical but loss sensitive, such as streaming multimedia and business-critical traffic. A typical use is for business applications subject to some form of reservation or admission control, such as capacity reservation per flow [14].

- **Excellent effort (3):** The UP associated with this traffic class represents traffic which can tolerate delay but is sensitive to loss. The delivery service for this traffic is deemed to be excellent (in terms of reliability of connection).

- **Best effort (2)**: The traffic associate with this class is delay and loss insensitive. In terms of prioritization, this class is provided with normal delivery service which is best possible at the prevailing network conditions.

- **Background (0):** This is the lowest priority traffic class characterized by traffic which is delay and loss insensitive. The purpose of creating a lower priority traffic class than the 'Best Effort' is to differentiate bulk transfers over networks so that their impact is minimized on the traffic generated users and applications.

At the IEEE 802.11e MAC, the UPs are mapped to four *access categories* (ACs). The UP-to-AC mapping given in [5] is shown in the Table 2-2 and a reference implementation model, highlighting the use of UP-to-AC mapping and per-AC EDCAF, as specified in [5], is given in the Figure 2-7.

Table 2-2: Higher layers UPs to MAC specific ACs mapping.

Priority	UP (Same as 802.1D user priority)	802.1D designation	AC	Designation (informative)
Lowest	1	BK	AC_BK	Background
	2	—	AC_BK	Background
	0	BE	AC_BE	Best Effort
	3	EE	AC_BE	Best Effort
	4	CL	AC_VI	Video
Highest	5	VI	AC_VI	Video
	6	VO	AC_VO	Voice
	7	NC	AC_VO	Voice

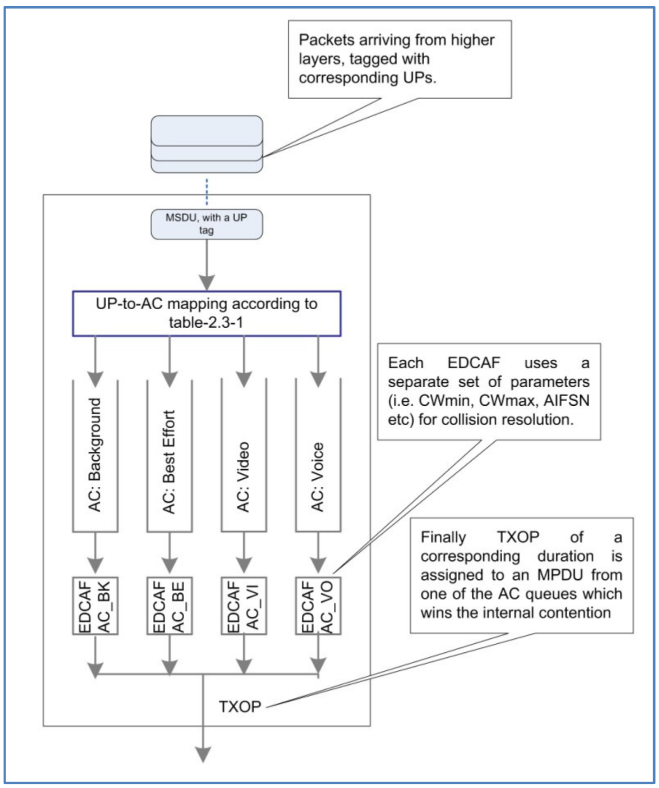

Figure 2-7: Mapping of UPs to corresponding MAC specific ACs' queues, resolution of inter AC contention. A reference implementation model as given in the IEEE 802.11 standard.

Each AC, uses an *enhanced distributed channel access function (EDCAF)* for contending for TXOPs. The EDCAF uses various parameters specific to an AC e.g. AIFS[AC], CW_{min}[AC], CW_{max}[AC], $TXOP_{limit}$[AC]; the values of these parameters are either obtained from the *EDCA* *parameter set* element in the beacon frames or uses default values in case when no EDCA parameter set is received from an AP. The AIFS duration used by an EDCAF for each AC is different and is given as:

$$AIFS[AC]=AIFSN[AC] \times aSlotTime + SIFS \tag{2.8}$$

The AIFSN[AC] is equal or greater than 2 for non-AP QoS STAs, e.g. the default values used by the non-AP STAs for AIFSN are 7, 3, 2, 2 for AC_BK, AC_BE, AC_VI and AC_VO respectively. The AIFSN[AC] for the AP should be greater or equal to 1. The default EDCA values for a QoS STA for AIFSN[AC] are given in Table 2-3, where as the values used by QoS AP which are advertised in beacons are given in Table 2-4 and values of aSlotTime and aSIFS for various PHY are given in Table 2-5. Using these values, the AIFS interval for various ACs using a specific PHY can be obtained. For instance the AIFS[VO], for PHY of the type ERP would be:

$$AIFSN[VO]= (2 \times 9)+ 10 = 28\ \mu s \tag{2.9}$$

For obtaining a TXOP, a STA uses the CS mechanism to find the status of medium occupancy. If the medium is idle, the EDCAF for each AC waits for AIFS[AC] interval and then waits further for the number of time slots as indicated, by the backoff counter. In EDCA the CW_{min} is different for each AC. An EDCAF after winning through the contention obtains a TXOP. In a TXOP, multiple frames belonging to the same AC for which the corresponding EDCAF obtains the TXOP can be transmitted. Frames belonging to other ACs cannot be transmitted within that TXOP. Due to the presence of four ACs, there is likelihood that two EDCAFs would win a TXOP, this is the case of *internal collision.* The internal collision is resolved in such a way that the lower priority AC gives in the TXOP to the higher priority AC and updates the CW[AC] while performing the next backoff. The backoff procedure is invoked in a similar fashion as it is done for DCF when a frame transmission fails. However in the case of EDCA, the retransmission counts (long and short) remain unchanged.

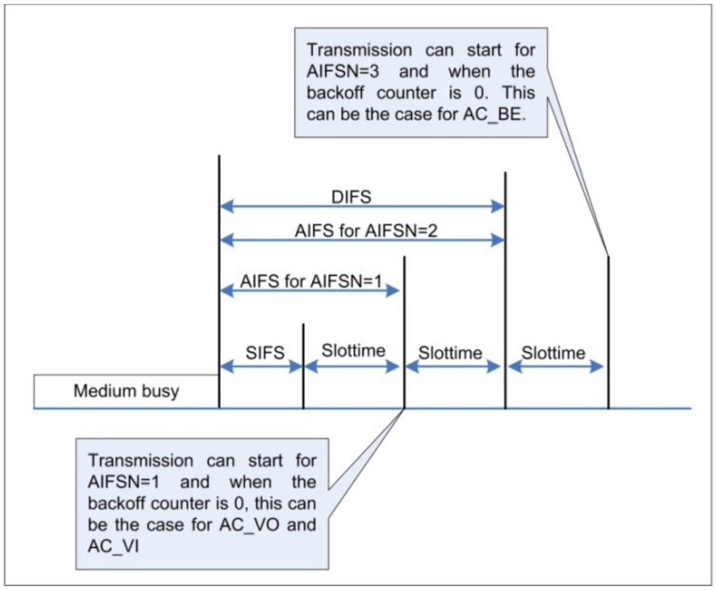

Figure 2-8: EDCA timing relationship. Notice the time difference in medium access among the four access categories. The AC_BK has to wait for 7 slottimes + SIFS while on the other hand AC_VO and AC_VI has to wait only for SIFS+Slottime.

Table 2-3: EDCA default parameters for non-AP QoS STA. In case of an ad-hoc network, STAs uses the default values for the EDCA parameters. It is important to notice that medium access time is prioritized for high priority ACs with the help of parameterized MAC. These parameters in turn depend on the underlying PHY.

EDCA parameters	Access categories			
	AC_BK	AC_BE	AC_VI	AC_VO
CWmin (0 to 255)	aCWmin[6]	aCWmin	(aCWmin - 1)/2 -1	(aCWmin - 1)/4 -1
CWmax (0 to 65535)	aCWmax	aCWmax	aCWmin	(aCWmin - 1)/2 -1
AIFSN (2 to 15)	7	3	2	2
TXOPLimit (0 to 65535)	0	0	3008 µs for OFDM (5 GHz) and ERP PHY, and 6016 µs for HR/DSSS PHY	1504 µs for OFDM (5 GHz) and ERP PHY, and 3264 µs for HR/DSSS PHY
MSDULifetime	500 TUs	500 TUs	500 TUs	500 TUs

[6] The contention window limits i.e. the aCWmin and aCWmax, are not fixed per PHY as the DCF. The default values per PHY of these two limiting parameters are given in Table-2.5.

Table 2-4: EDCA default parameters at QoS AP which are advertised to the STAs in BSS in beacons.

EDCA parameters	Access categories			
	AC_BK	AC_BE	AC_VI	AC_VO
CWmin (0 to 255)	aCWmin	aCWmin	(aCWmin - 1)/2 -1	(aCWmin - 1)/4 -1
CWmax (0 to 65535)	aCWmax	4 x(aCWmin + 1) - 1	aCWmin	(aCWmin - 1)/2 -1
AIFSN (2 to 15)	7	3	1	1
TXOPLimit (0 to 65535)	0	0	3008 µs for OFDM (5 GHz) and ERP PHY, and 6016 µs for HR/DSSS PHY	1504 µs for OFDM (5 GHz) and ERP PHY, and 3264 µs for HR/DSSS PHY
MSDULifetime	500 TUs	500 TUs	500 TUs	500 TUs

EDCA admission control.

An admission control scheme for EDCA is suggested in the standard. Under the EDCA admission control, each AC informs the QAP about the *traffic specification (TSPEC)*. The TSEPC includes several parameters e.g. the mean/peak data-rate, mean/peak frame size etc (See annex-k given in [5] for detail of TSPEC elements and TSPEC construction). Every QoS STA maintains two variables called the *admitted_time* and the *used_time.* The HC calculates the amount of time that the AC can use to access the medium (called the *medium_time*) and sends this information to the QoS STA. The STA updates the *admitted_time* variable to this value in the response frame. Every time a STA access the medium, it updates the *used_time* for the corresponding AC. An AC is not allowed to transmit beyond the *admitted_time*, and if it requires more time than the *admitted_time* it has to generate a new request.

2.4.8.2 HCCA

In the HCCA mechanism an HC manages transmission of QoS traffic from the AP to QoS STAs and polling of non-AP STAs. The HC essentially works as a PC in a BSS and uses CFP for delivery of frames to non-AP STAs however it is not mandatory for the HC to use CFP. The QAP can initiate several CF bursts between a beacon interval, such CF bursts are called *controlled access periods (CAPs)*. The HCCA provides flexibility as compared to PCF in a way that CAPs can be generated at any time when the HC detects the medium to be idle for PIFS. As a result of the flexibility of HCCA and its capability to do the job of PCF, the PCF is left redundant and therefore it is supported but its use is optional. The HCCA also provides the use of EDCA within a beacon interval by putting a limit on the use of CAPs (in terms of time); this limiting value is specified by a variable called $T_{CAPlimit}$.

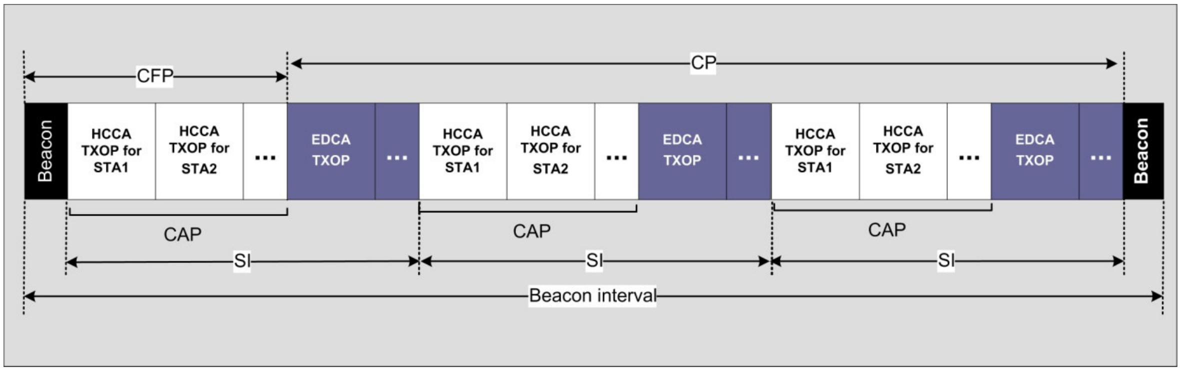

Figure 2-9: HCCA operation: allocation of HCCA TXOPs , generation of CFP just after the beacon, then several CFP bursts throughout the beacon interval called CAPs followed by EDCA (contention based) TXOPs.

Admission control in HCCA

An admission control strategy improves the performance of QoS BSS, the standard does specify a reference model for accomplishing admission control, however such algorithms are left open. Admission control strategies depend on the outcome of a scheduling algorithm. According to the reference implementation model, a QSTA sends an admission request for a *traffic stream (TS)* to the QoS AP before the transmission of any frames from an AC. A QoS STA can admit eight TSs with different priorities. The TSs has the same meaning when HCCA is operational as the ACs have with EDCA mode of the IEEE 802.11e is in operation. In the request for admission of a TS, a QoS STA sends TSPECs which includes specification for mean/peak data-rate, mean/peak frame size, delay bound and maximum *required service interval (RSI)*. The maximum RSI specifies the delay tolerance of an application. The HC after receiving an admission request, determines the service interval (SI), which should be the highest sub-multiple of the beacon interval and should not be longer than the highest RSI for all admitted/requested TSs. After selecting an appropriate SI according to the requested credentials of all TSs, the HC divides the beacon into integer number of SIs and various STAs are polled in every SI. Each STA is assigned TXOP according to the request. This process is explained with the help of Figure 2-9.

The reference admission control unit takes a decision by using a very simple inequality i.e. for assigning TXOP to a $(K+1)^{th}$ TS, the admission control unit has to make it sure that the combined sum of all TXOPs within every SI is less than or equal to the maximum CAP duration within the beacon interval. In a simple form, this in equality can be written as given in [5]:

$$\frac{TXOP_{(k+1)}}{SI} + \sum_{i=1}^{k} \frac{TXOP_i}{SI} \leq \frac{T_{CAPlimit}}{T_{Beacon}} \qquad (2.10)$$

2.5 IEEE 802.11 PHY Layer

The IEEE 802.11 standard defines various PHYs. Each PHY embodies two protocol functions:

1. PHY media dependent (PMD) system, which defines the characteristics and method of transmitting and receiving frames over the wireless medium, e.g. the type of modulation, the number of channels, the transmission power etc.

2. PHY layer convergence protocol (PLCP), which defines a method of mapping the IEEE 802.11 MPDUs into the frame formats suitable for transmission and reception between two STAs under the associated PMD system. A PLCP is unique to a PMD and as a result the PLCP mapping mechanism varies with the variation in the PMD characteristics.

A reference model showing the interaction and hierarchy of various PHY-sublayers with the MAC is given in Figure 2-10.

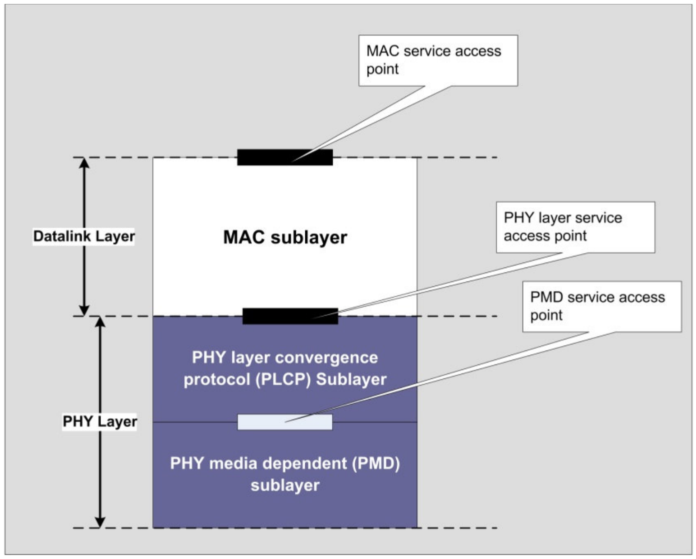

Figure 2-10: Reference model for highlighting the hierarchy and interaction of various PHY-sub-layers with the MAC and higher layers as defined in the IEEE 802.11 standard.

In order to transmit PHY service data units (PSDUs) to STAs within a BSS, a PLCP preamble and a PCLP-header are attached to the PSDUs to from PHY protocol data units (PPDUs). The PLCP preamble allows the PHY to establish synchronisation and define the start of frame. The PLCP header contains information about the length of whitened PSDU, PLCP signalling field for conveying the transmission rate of the PSDU and header checksum field. The format of long-PPDU as defined under the IEEE 802.11b, [3], is given in Figure 2-11.

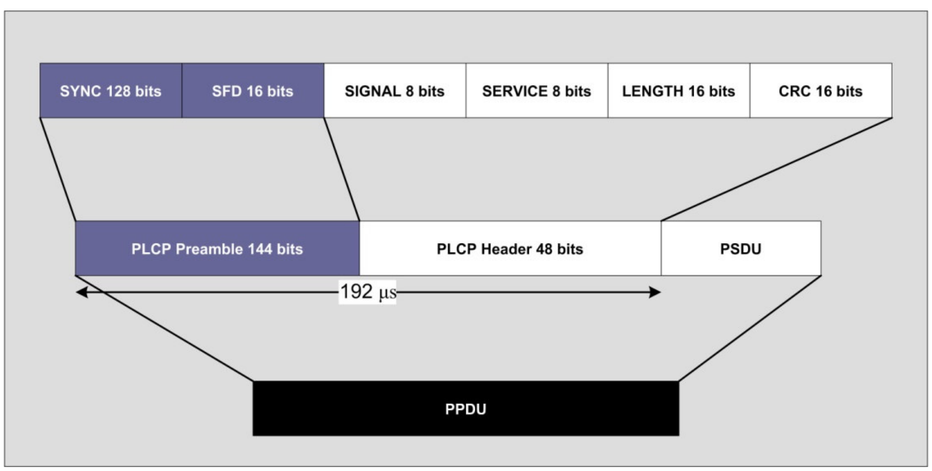

Figure 2-11: PPDU format.

The PLCP preamble and PLCP header are transmitted at 1Mbps while the PSDU part of the PPDU can be transmitted at any of the supported transmission rates.

The high rate DSSS (HR/DSSS) is an extension of the DSSS system and provides support of 5.5 Mbps and 11 Mbps payload transmission data-rates. The HR/DSSS also provides an optional mode while operating at 2, 5.5 or 11 Mbps for increasing throughput by using a shorter PLCP preamble. To interoperate with a receiver which is not capable of receiving short-preamble (i.e. the DSSS) the transmitter uses the long preamble and header.

The Extended rate PHY (ERP) builds on the specification of OFDM PHY (for IEEE 802.11a 5 GHz) to provide payload transmission support at 6, 9, 12, 18, 24, 36, 48 and 54 Mbps. Two additional ERP-PBCC modulation modes are defined to support payload transmission at 22 and 33 Mbps. An optional modulation mode called DSSS-OFDM is also defined to support payload transmission at 6, 9, 12, 18, 24, 36, 48 and 54 Mbps.

A summary of PHY specific and MIB parameters and their typical values is given in Table 2-5.

Table 2-5: PHY (PLCP) specific and MIB parameters values (OFDM PHY operating in the 5 GHz band is omitted from this table)

PHY	Supported data rates in Mbps	Modulation	MPDU (PSDU) maximum Length	Slot time	SIFS	Preamble Length	PLCP Header Length	CWmin	CWmax
FHSS	1, 2	2GFSK for 1 Mbps and 4GFSK for 2 Mbps	4095 octets (The recommended value of Max MPDU for FHSS is 400 and 800 octets for 1 and 2 Mbps respectively, which corresponds to the frame duration of 3.5ms [reference: 802.11 standard]	50 µs	28 µs	96 µs (96 bits)	32 µs (32 bits)	15	1023
DSSS	1, 2	DBPSK for 1 Mbps and DQPSK for 2 Mbps		20 µs	10 µs	144 µs	48 µs	31	1023
IR	1, 2	16-PPM for 1 Mbps and 4-PPm for 2 Mbps	2500 Octets	8 µs	10 µs	16 µs (1Mbps) 20 µs (2Mbps)	41 µs (1 Mbps) 25 µs (2 Mbps)	63	1023
HR/DSSS	1, 2, 5.5, 11	DBPSK for 1 Mbps, DQPSK for 2 Mbps, CCK for 5.5 and 11 Mbps		20 µs	10 µs	144 µs (Long Preamble) 72µs (Short preamble, 72 bits at 1 Mbps)	48 µs (Long Preamble) 24µs (short preamble, 48 bits at 2 Mbps)	31	1023
ERP	1, 2, 5.5, 6, 9, 11, 12, 18, 24, 36, 48, and 54	DBPSK for 1 Mbps, DQPSK for 2 Mbps, CCK for 5.5 and 11 Mbps, BPSK (6Mbps), BPSK (9Mbps), QPSK (12, 18 Mbps), 16-QAM (24,36 Mbps), 64-QAM (48, 54 Mbps)	4095 octets	20 µs or 9 µs when BSS operates under ERP	10 µs	144 µs (Long Preamble) 72µs (short preamble of DSSS-OFDM),	48 µs (Long Preamble) 24µs (short preamble of DSSS-OFDM)	CWmin (0) = 31, CWmin (1)= 15	1023

2.6 Multi-rate support.

STAs which uses IEEE 802.11 conformant MAC and PHY have the ability to operate at various transmission rates. Varying the transmission rate according to the variations in the characteristics of wireless medium can potentially improve the overall performance. The existing standard, including the approved extensions, does not provide specification for a rate-adaptation scheme. The currently active draft of IEEE 802.11n provides specification of a rate-adaptation scheme (please see section-3.2 of the chapter-3). Although there is no standard specification for a rate-adaptation strategy, the IEEE 802.11 standard gives certain mandatory rules for devising such a strategy. But before going into the detail of various rules which STAs should follow while using a rate-adaptation scheme, it is important to reiterate the mechanism through which STAs communicate their multi-rate capability information across the BSS.

In a BSS, a STA, which starts the BSS, designates a set of transmission rates, which should be supported by all STAs in a BSS; this set is called *BSSBasicRateSet*. A STA communicates its multi-rate capability in the form of a parameter called *Supported rates*. The *Supported rates* parameter is also included in association request, re-association request and probe request frames. The *Supported rates* parameter includes all the operational rates at which a STA can transmit and receive. The *supported rates* parameter is a superset of rates represented in the BSSBasicRateSet. This parameter can hold information for only eight operational rates; for STAs which support more than eight operational rates an *Extended supported rates* parameter is used in all relevant frames (such as the association/re-association/probe request frames). The *supported rate* parameter is encoded as one to eight octets, where each octet represents an operational rate. For an operational rate which is also a part of the BSSBasicRateSet, the first bit of the octet is 1, while the rest of 7 bits are used to encode the operational rate according to the Table 2-6. For example, for an operational rate of 2 Mbps which if a part of the BSSBasicRateSet, the octet representing this rate would be 1000 0100 (132 decimal). Likewise, as an example, for operational rate of 6 Mbps, which if not included in the BSSBasicRateSet would be encoded as 0000 1100 (12 decimal).

Therefore, management frames (beacon, association response, re-association response and probe response) include the *supported rates* parameter to convey the operational rates and the BSSBasicRateSet to STAs. STAs also convey their operational rates to an AP by including the *supported* rates and the optional *extended supported rates* in various management frames such as the association/re-association and probe request frames. Association can be denied within a BSS if the OperationalRateSet of a station does not include the transmission rates included in the BSSBasicRateSet.

Table 2-6: Encoding transmission rates to be used in the *supported rate parameter*.

Value (in decimal)	Transmission rates
2	1 Mbps
3	1.5 Mbps
4	2 Mbps
5	2.5 Mbps
6	3 Mbps
9	4.5 Mbps
11	5.5 Mbps
12	6 Mbps
18	9 Mbps
22	11 Mbps
24	12 Mbps
27	13.5 Mbps
36	18 Mbps
44	22 Mbps
48	24 Mbps
54	27 Mbps
66	33 Mbps
72	36 Mbps
96	48 Mbps
108	54 Mbps

2.6.1 Mandatory rules for a rate-adaptation scheme

Protection mechanism frames (RTS/CTS, CTS-to-self), whose purpose is to propagate medium usage information across the BSS and to establish the virtual CS mechanism should be transmitted at such a rate so that ERP and non-ERP STAs can interpret them and know about the duration of medium usage[7]. For this purpose the standard mandates that such frames should only be sent at one of the mandatory transmission rates of the DSSS PHY (IEEE 802.11), or the HR/DSSS (IEEE 802.11b) so that all STAs within a BSA can decode the transmission and update their corresponding NAVs for the duration of transmission.

[7] In a BSS where ERP and non-ERP STAs coexist and the ERP uses either DSSS-OFDM or ERP-PBCC preamble then in this case there is no need to use the protection mechanism frames as the DSSS header.

With an exception of the frame types mentioned in the above paragraph and *Block acknowledgment request/Block acknowledgment* frames, all other control frames should be sent at one of the rates in the BSSBasicRateSet.

Broadcast/multicast frames should also be sent at one the rates in the BSSBasicRateSet.

Frames for polling stations (CF-Poll) generated within a CP should only be sent at the one of the rates in the BSSBasicRateSet, this condition is not required if the protection mechanism (e.g. the RTS/CTS) is used before generating the CF-Poll.

In situation when the supported rate set of the receiving STA is not known, the transmitting STA should only use rates specified in the BSSBasicRateSet or should transmit at a rate at which it received frames from the receiving STAs.

In a normal frame exchange, a transmitting STA needs to inform other STAs in a BSS about the duration of medium usage in a frame exchange by using the *duration* field in frames. This value in the duration field includes the time for transmission of a frame from the transmitting STA, the inter-frame spaces, and the time that a corresponding ACK frame would take. In this process, a transmitting STA knows about everything apart from the duration of ACK/CTS; because it is up to the receiving STA to select a certain transmission rate for sending the ACK/CTS frame. Therefore, to enable the transmitting STA to calculate the value of the *duration* field, the standard mandates that a receiving STA (which would send an ACK/CTS frame) should send an ACK/CTS frame at the highest rate in the BSSBasicRateSet which is less than or equal to the rate at which the transmitting STA sent the latest frame in the frame exchange sequence (which can be RTS or data frame).

However, if the transmitting STA sent the frame which is not in the BSSBasicRateSet (and thus the condition outlined in the above paragraph could not be met), then a receiving STA should send the ACK/CTS frame at the highest mandatory rate of the PHY which is less than or equal to the rate at which the transmitting STA sent a frame.

The rules mandated by the standard for devising a rate-adaptation strategy essentially means that transmissions should only be done at rates which could be successfully decoded at the receiver, and at the same time the selection of transmission rates should not disrupt the distributed MAC operations in the BSS.

2.7 Summary

This chapter gives a detailed insight in the IEEE 802.11 family of standards. The objective is to highlight the specific detail of MAC and PHY layer characteristics defined in the standard, which are referred to in the later chapters. The operational characteristics of various coordination functions are explained which forms the basis of research schemes and the related discussion in chapter-3 and chapter-5. The standard does not specify a rate-adaptation mechanism; however, it specifies a comprehensive set of mandatory rules for designing a rate-adaptation mechanism. These rules have been discussed in the perspective of designing a rate-adaptation scheme.

Conceptually, Chapter-2 forms the basis for understanding the key contributions of this thesis.

2.8 References:

[1] "IEEE Std 802.11-1997 Information Technology- telecommunications And Information exchange Between Systems-Local And Metropolitan Area Networks-specific Requirements-part 11: Wireless LAN Medium Access Control (MAC) And Physical Layer (PHY) Specifications," IEEE Std 802.11-1997 , vol., no., pp.i-445, 18 Nov 1997.

[2] "Supplement to IEEE standard for information technology telecommunications and information exchange between systems - local and metropolitan area networks - specific requirements. Part 11: wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: high-speed physical layer in the 5 GHz band," IEEE Std 802.11a-1999 , vol., no., pp.-, 1999.

[3] "Supplement To IEEE Standard For Information Technology- Telecommunications And Information Exchange Between Systems- Local And Metropolitan Area Networks- Specific Requirements- Part 11: Wireless LAN Medium Access Control (MAC) And Physical Layer (PHY) Specifications: Higher-speed Physical Layer Extension In The 2.4 GHz Band," IEEE Std 802.11b-1999 , vol., no., pp.i-90, 2000.

[4] "IEEE standard for information technology - telecommunications and information exchange between systems - local and metropolitan area networks - specific requirement. Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specification. Amendment 3: specifications for operation in additional regulatory domains," IEEE Std 802.11d-2001 , vol., no., pp.-, 2001.

[5] "IEEE Standard for Information technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," IEEE Std 802.11-2007 (Revision of IEEE Std 802.11-1999) , vol., no., pp.C1-1184, June 12 2007.

[6] "IEEE Std. 802.11h - 2003," IEEE Std 802.11h-2003 (Amendment to IEEE Std 802.11, 1999 Edn. (Reaff 2003)) , vol., no., pp.0_1-59, 2003.

[7] "IEEE Standard for Information technology- Telecommunications and information exchange between systems- Local and metropolitan area networks- Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer

(PHY) specifications Amendment 6: Medium Access Control (MAC) Security Enhancements," IEEE Std 802.11i-2004 , vol., no., pp. 0_1-175, 2004.

[8] "IEEE Standard for Information technology- Telecommunications and information exchange between systems- Local and metropolitan area networks- Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 7: 4.9 GHz-5 GHz Operation in Japan," IEEE Std 802.11j-2004 , vol., no., pp. 0_1-40, 2004.

[9] "IEEE Standard for information technology- Telecommunications and information exchange between systems- Local and metropolitan area networks- Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 1: Radio Resource Measurement of Wireless LANs," IEEE Std 802.11k-2008 (Amendment to IEEE Std 802.11-2007) , vol., no., pp.c1-222, June 12 2008.

[10] "IEEE standard for information technology-telecommunications and information exchange between systems-local and metropolitan area networks-specific requirements part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 2: fast basic service set (BSS)," IEEE Std 802.11r-2008 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008) , vol., no., pp.c1-108, July 15 2008.

[11] "IEEE Draft STANDARD for Information Technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Amendment 4: Enhancements for Higher Throughput," IEEE Unapproved Draft Std P802.11n/D4.00, Mar 2008 , vol., no., pp.-, 2008.

[12] "Draft Amendment to Standard for Information Technology- Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Amendment 7: Interworking with External Networks," IEEE Unapproved Draft Std P802.11u/D3.0, May 2008 , vol., no., pp.-, 2008.

[13] "IEEE standard for Local and metropolitan area networks Media Access Control (MAC) Bridges," *IEEE Std. 802.1D-2004 (Revision of IEEE Std. 802.1D-1998),* pp. 1-269, 2004.

[14] W. Stallings, "LAN QoS", The Internet Protocol Journal, vol-4, issue-1. Available online: http://www.cisco.com/web/about/ac123/ac147/archived_issues/ipj_4-1/lan_qos.html

[15] W. Stallings, "Local and Metropolitan Area Networks", Sixth Edition, Prentice Hall, 2000.

Chapter 3

Design of a Rate-adaptation protocol for multi-rate capable PHY

3.1 Introduction and motivation

The IEEE 802.11 standard [1] conformant wireless communication devices have the ability to communicate at multiple transmission rates. The lowest transmission rate supported by the standard is 1 Mbps while the highest transmission rate (in the approved extensions of the standard) is 54 Mbps.

Transmission rate holds a great significance in determining the overall efficiency of a communication system. Generally, it is always desirable to use the highest transmission rate e.g., highest transmission rate can potentially yield: highest throughput, lower medium occupancy (thus lower contention delays and higher spectrum efficiency) and lower power consumption. The IEEE 802.11 standard defines a receiver's minimum input level sensitivity for various transmission rates. The standard specifies an upper limit of the frame error rate (FER); a receiver should experience less than 10% FER when receiving a PSDU of length 1000 octets at each of the specified levels of signal strength (please see section-3.13.3 for detail of various reception and sensitivity thresholds). Because of the underlying modulation schemes the sensitivity threshold increases for transmissions made at a higher rate. Therefore, in situations when the SNR fluctuates (due to a variety of reasons) the higher rate transmissions are prone to erroneous reception as compared to lower rate transmissions.

Inappropriate rate selection can put a communication system in either of two states: (i) a transmitter operates at lower than optimum rate, thus the chances of transmission failures (and thus retransmissions) are minimized, or (ii) a transmitter uses an inappropriately selected higher rate, which minimizes the chances of successful transmissions.

A low-rate station in a Basic Service Set (BSS) causes rate-anomaly [2, 3]. Rate-anomaly occurs when low-rate and high-rate co-exist in the same BSS and the low-rate stations penalize the performance of high-rate stations by occupying the medium for unnecessarily longer durations. On the other hand always using the highest transmission rate increases the probability of transmission failures. Every communication system pays a high price when it suffers from transmission failures. In case of the IEEE 802.11 standard conformant stations, a station after sending a frame waits for '*ACK-timeout*' duration till it concludes that frame has failed to reach the destination. After a transmission failure, the standard MAC, increments the backoff period, which essentially means that the station would have to wait for a longer time before accessing the medium for retransmitting the frame. The '*Ack-Timeout'* is equal to:

$$\textit{Ack-Timeout} = \text{SIFS time} + \text{SlotTime} + \text{PHY-RX-START-Delay} \qquad (3.1)$$

Moreover, selection of an inappropriate transmission-rate not only misuses the shared medium and cause delays, it also causes overhead in terms of energy, in the battery powered mobile terminals [4].

Ideally, for efficient communication, a station should reduce the transmission rate when its receiver is unable to receive and should increase the transmission rate whenever possible. This process of selection of appropriate transmission rate according to prevailing medium/communication status is called rate-adaptation.

Although, rate-adaptation protocol is of significant importance, the IEEE 802.11 standard does not provide standard specification for a rate-adaptation technique. The standard, however, specified mandatory rules for devising a rate adaptation technique (please see chapter-2 section-2.6.1 for detail of the mandatory standard rules for rate-adaptation schemes). As a result of the lack of standard specification, there are a number of techniques proposed by manufacturers of the Standard conformant devices and from independent researchers. Each of the proposed approaches addresses the issue of rate-adaptation from a different perspective, keeping in view the underlying timing constraints of the MAC sub-layer.

This chapter gives a detailed insight into various rate-adaptation techniques, analyses various aspects of such protocols and focuses on their implications on the performance of communication.

In this chapter, we present our proposed rate-adaptation solutions:

4. A frame-failure-statistics based rate adaptation solution which uses an *on-demand incremental* strategy for selecting a rate-selection threshold. This solution is based on a cross-layer communication framework, where the rate-adaptation module involves information to/from the Application layer along with relevant information from the MAC sub-layer while making a decision.
    a. The on-demand incremental strategy avoids the chances of retransmissions.
    b. The rate-control function works in close coordination with the Application layer requirements.
    c. The underlying MAC sub-layer's timing constraints and medium congestion is taken into account while performing rate-adaptation.
    d. A feedback mechanism is used to communicate the underlying layers' status and capability information to higher layers.
    e. A novel mechanism of frame-loss differentiation is introduced to distinguish between frames lost because of erroneous transmissions (lost frames) from frames which are not received because of simultaneous transmission by other stations.
5. An SNR-based rate-adaptation scheme, called MutFed, relying on mutual feedback between a transmitter and receiver pair.
    a. MutFed uses a novel mechanism of conveying feedback between communication peers, without the need for changing the Standard frame formats.
    b. The proposed rate-adaptation scheme is highly responsive to variations in communication-channel's quality.
    c. MutFed uses an extremely efficient and highly responsive frame-loss differentiation mechanism for distinguishing *corrupted frames* from *collided frames.*

Various features associated generally with all rate-adaptation schemes, and specifically with the proposed rate-adaptation mechanisms are discussed in detail in this chapter. The designs presented in this chapter are elaborated further with the help of performance tests and analysis in the following chapter (4). This chapter also gives a detailed insight into various MAC specific issues and forms the basis for the research presented in chapter-5.

3.2 Review of existing rate adaptation protocols

A rate-adaptation scheme monitors the *channel state information* (CSI) and accordingly reacts by selecting an appropriate transmission rate. The methods of obtaining the CSI are broadly categorized into two categories: (i) Estimation of CSI through the transmission history (success/failures) of previous frame transmissions at the transmitting station; and (ii) estimation of CSI through the received signal strength and noise at the receiver side, and providing the feedback to the transmitter.

All the proposed rate-adaptation schemes use either or both of the above mentioned methods of obtaining the CSI while performing the rate-adaptation.

3.2.1 Rate-adaptation and CSI estimation through frame success/failure

The frame transmission statistics provide an estimation of the channel quality. Generally, when the number of successfully transmitted frames is higher in a particular estimation window, it reflects the possibility of a future successful transmission i.e.

$$P_{success}^{R} = \frac{\mu_R}{W}$$

Where, $P_{success}^{R}$ is the probability of successful transmission at the transmission rate '*R*', μ_R is the number of successfully transmitted frames at '*R*' (so far, in the current estimation window), and the parameter '*W*' represents the size of the estimation window. Therefore, once the probability of success/failure of frame transmission at a particular transmission rate is known, the channel quality can be estimated. Such estimations form the basis of *statistics-based* rate adaptation schemes.

In the statistics based rate-adaptation schemes, the channel estimation and rate-selection are performed by the sender and therefore such schemes are also called *sender-side* rate-adaptation schemes.

Generally, if the size of the *estimation window (W)* is large, it would filter out short-term channel variations. While on the other hand, keeping the estimation-window of a smaller size, would force the rate-adaptation scheme to over-react to short-term channel variations.

3.2.1.1 Auto Rate Fallback (ARF)

It is commonly agreed that the first known rate-adaptation scheme for IEEE 802.11 conformant wireless communication devices was published in [5] for Wave-LAN®-II in 1997. The rate-adaptation scheme was called Auto Rate Fallback (ARF) was introduced for Wav-LAN-II devices and it is one of the highly cited publications on rate-adaptation. The Wave-LAN-II supported transmission at 1 Mbps and 2Mbps. According to [5] the different modulation techniques used for different transmission-rates of Wave-LAN-II indicated that transmissions at lower rates would be more robust than transmission at higher rates. Therefore, transmission rates were mapped (associated) to transmission ranges by introducing the idea of *inner circle* for 2Mbps and *outer circle* for 1Mbps transmission rates, shown in the Figure 3-1. A station moving in the outer circle would be able to successfully transmit at 1 Mbps however, for communication efficiency when the station moves into the inner circle, it should automatically switch to a higher transmission rate of 2 Mbps. This automatic selection of transmission rates was handled by ARF.

The ARF scheme keeps track of a timing function and uses statistics related to the status of most recent frame transmissions. The default transmission rate selected by the ARF is 2 Mbps. ARF lowers the transmission rate to 1 Mbps after 2 consecutive frame failures, indicated by missing ACK frames at the transmitter. When an ACK is missed for the first time following an earlier successful transmission, the first retry is made at 2 Mbps. In case of another, consecutive frame loss, marked by missing ACK the ARF lowers the transmission rate to 1 Mbps and all the subsequent retries and frame transmissions are performed at 1 Mbps.

After lowering the transmission rate a timer is started to track successfully transmitted frames and lost frames. When either the timer expires or the number of successfully transmitted frames reaches 10, ARF increases the transmission rate back to higher rate (2 Mbps). The next frame transmission, which can be called a probe transmission, is performed at the higher rate. If the transmission at the higher rate (when using the probe transmission) fails, the ARF immediately lowers the transmission rate again to 1 Mbps. This process is repeated in a similar fashion and ARF attempts transmission at higher rates after 10 successful transmissions or timer expiry.

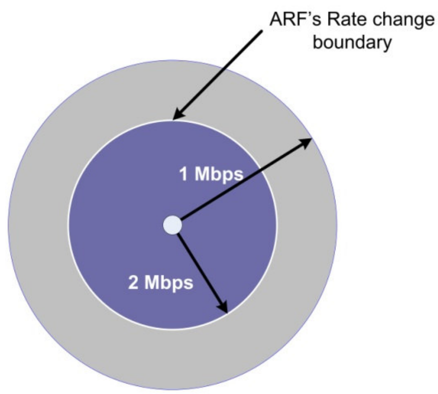

Figure 3-1: ARF's rate change boundary as given in [5].

3.2.1.2 Adaptive-ARF (AARF)

ARF responds very well to short term channel variation by reducing the transmission rate after two successive frame losses. However, in stable conditions or long term variations it does not perform well because it inherently tries to increase the transmission rate after every 10 successful transmissions. The authors in [6] proposed Adaptive ARF (AARF), by introducing binary exponential backoff to ARF's higher threshold.

According to AARF, if the first rate-up attempt after 10 successful transmissions fails, then the next rate-up attempt should be made after 20 frames. If the situation persist, this threshold should be doubled every time; the highest limit that a threshold can reach as a result of the BEB is set to 50. For example, assume that a station successfully transmitted 10 frames at 5.5 Mbps, according to the ARF, it would send the next frame at a higher rate and if it fails it should again rate-down to 5.5 Mbps and repeat this process after next 10 successful frames. While in the case of AARF, the station would repeat the rate-up after 20 successful frame transmissions. AARF essentially minimizes the frequency of rate-up attempts as proposed for ARF. AARF would logically perform well in long term channel variations, because it would not attempt higher-rate transmissions and thus avoid frame failures. However, at the same time, it slows down its responsiveness variations in the channel conditions.

3.2.1.3 SampleRate

The ARF and AARF react sequentially to improvements in the channel quality. In order to improve the responsiveness, SampleRate [7], randomly selects transmission rates. SampleRate sends frames at a transmission rate which would provide the highest throughput. Initially when the transmitter starts sending the frames, it uses the highest possible transmission-rate value. It stops transmitting at a rate if it experiences 4 successive losses. It will keep on decreasing the transmission-rate until it finds a rate value which can successfully transmit frames. Every 10^{th} frame, it randomly selects a transmission-rate value which it believes would provide better throughput than the current one. It does not try a transmission-rate value if (operating at) that value causes 4 successive frame losses or if it's lossless transmission time is more than the average transmission time of the current data-rate value. SampleRate uses a 10 seconds window for calculating the average transmission time.

3.2.1.4 Multi-Rate Retry (MRR) of MADWIFI

Multi-rate retry (MRR) is a rate-adaptation scheme used in the Multiband Atheros Driver for WiFi (MADWIFI) [8]. MADWIFI is a Linux driver for Atheros based chipsets used in the IEEE 802.11 standard devices. MADWiFi is a semi-open-source driver. It hides the hardware specific functionalities using a binary only Hardware Abstraction Layer (HAL). Transmissions are handled with *transmission descriptors* (which is a structure *ath_desc*). The transmission descriptor structure defines a *union* (called ds_us) having two structures, one indicating the transmission status and another for receiving status. It also has a pointer to the next descriptor and to the data buffer which is to be transmitted and an ordered set of 4 pairs of rate and transmission count fields (r_0/c_0, r_1/c_1, r_2/c_2 and r_3/c_3). The transmission status structure also indicates the transmission rate at which the frame was transmitted.

According to the MRR scheme, the frame transmission starts at r_0 but if transmissions fail c_0 times the rate is lowered to r_1 and c_1 attempts are made at this rate. This process is repeated till either a frame is successfully transmitted or when transmission at r_3 is repeated c_3 times, after which the packet is discarded and the transmission status is updated in the descriptor. The driver periodically changes the value of r_0/c_0, r_1/c_1, r_2/c_2 and r_3/c_3 according to the transmission status. This periodic duration is 0.5 to 1 second. Therefore, short term channel variations are handled by switching from r_0 to r_1 to r_2 and finally to r_3, while for long term variations it periodically updates the values of r_0/c_0, r_1/c_1, r_2/c_2 and r_3/c_3.

3.2.1.5 Adaptive Multi-rate retry (AMRR)

The authors in [6], introduces binary exponential backoff (BEB) procedure to the original MADWIFI rate-adaptation mechanism. This is called Adaptive MRR (AMRR). The idea behind AMRR is that it adaptively changes the length of the period after which (that period of time) the values of the rate/count pairs (used for rate adaptation in the original MADWIFI driver) are changed. To ensure responsiveness of the algorithm to short-term channel variations AMRR uses c_0=1, c_1=1, c_2=1 and c_3=1 where as the MADWIFI used c_0=4, c_1=2, c_2=2 and c_3=2. The parameter 'r_3' is always set to the lowest available value of transmission-rate while 'r_1' and 'r_2' are set to consecutively lower rates than 'r_0'. AMRR increases the transmission-rate (in this case, updates the whole set of rates from r_0 to r_2) if less than 10% of transmitted frames fail during the previous period (it also checks if the information is not too old) and the success rate crosses a *success threshold*. On the other hand if 33% of transmitted frames fail during a previous period, it doubles the *success threshold* for next time (and that is how it introduces the BEB in the standard MADWIFI).

3.2.1.6 ONOE

Onoe [8] is a credit based algorithm and works by incrementing/decrementing credits for a particular transmission rate value if the frame loss percentage is lower/higher than 10 percent in a periodic fashion with a default period of 1 second. A higher transmission rate value is selected if the number of credits crosses a specific threshold (10 or more). If no frames are successfully transmitted the transmission rate is lowered to the lower available value. The credits are reset to zero each time a new transmission-rate value is selected.

3.2.1.7 Miscellaneous approaches

Opportunistic Auto Rate (OAR) [9, 10] protocol aims at exploiting the duration of high-quality channel conditions. The core idea of OAR is to send back-to-back data frames whenever the channel conditions are good. OAR works in coordination with another rate adaptation algorithm (RBAR and optionally with ARF). Primarily, the rate adaptation algorithms are responsible for determining a suitable transmission value, and the OAR then maintains the selected transmission rate for successive transmissions when the channel conditions are good. A similar approach is used in OSAR [11].

An enhancement to ARF is proposed in [12]. According to authors in [12] ARF doesn't reflect the current contention levels and as a result if the contention for medium usage increases, there are higher chances of frame losses. Their argument is based on the findings in [13] which has shown with the help of Markov chain analysis of ARF that the rate distribution of ARF is mainly distributed on low transmission rates when the number of contending stations increase. In their opinion, the fixed thresholds used by ARF are not optimum in such a scenario and ARF would decrease the transmission rate with increase in contention even if the frame error rate remains stable. Their design philosophy is that the probability of increasing and decreasing the transmission rate of a transmitting station should not depend on the number of contending stations if the channel error rate is stable. Therefore, according to their proposed scheme for rate adaptation, the *rate-up* and *rate-down* thresholds are updated every time the backoff counter reaches zero to include the effect of medium contention. However, simply using ARF with the proposed enhancements still suffers from the unnecessary back and forth rate selection; therefore, to minimize this effect they used further enhancement which is similar to AARF.

To avoid using fixed, predefined thresholds, a machine learning technique, stochastic learning automata, is applied to rate-adaptation in [14], to randomly select a rate for a transmission and dynamically update the decision based on the ACKs feedback.

A semi-Markovian framework for analyzing the performance of ARF and AARF is presented in [15], shows the neither of the two rate-adaptation schemes consistently outperforms each other in all conditions. While ARF responds relatively quickly to improved channel conditions when compared with AARF, the later is better in long term channel variations.

The authors in [16], analyzed various rate adaptation algorithms using a testbed in variety of test scenarios. According to their findings SampleRate and ONOE experience drop in frames delivery rate when the number of transmitters increase. One of the reasons of poor performance in the case of SampleRate is that in situations of higher node density, it frequently samples various available rates and then it has few samples for accurate estimation of transmission time. RSSI based algorithms however appear to perform better in case of collisions and report better cumulative throughput. ONOE and SampleRate can also show performance improvements with use of RTS/CTS mechanisms for minimizing losses due to hidden nodes.

Rate-adaptation schemes have been broadly placed in two categories according to their inherent design for rate-adaptation in [17]: throughput-based rate-adaptation schemes which include [18], [19] and [20], and error-based rate-adaptation schemes like [21] and [22]. The authors proposed the use of appropriate rate-adaptation scheme on per-frame basis according to the type of frame e.g. frames which are associated with throughput demanding applications should use throughput-based rate-adaptation while on the other hand loss sensitive applications should use rate-adaptation schemes which are cautious in terms of frame losses.

There are a number of approaches which have modified the existing rate-adaptation schemes (mainly the ARF), however, quite unfortunately the assumptions on which these schemes are based prevent the practical realization of such rate-adaptation schemes. For instance, [23] proposed a rate-adaptation scheme for downlink, where a station overhears the transmission from AP to other stations and if the AP uses higher transmission rates for other stations than this station, then this station provides a feedback to the AP for increasing the transmission rates. This scheme is based on a number of assumptions, for example, there has to be a number (which in itself is unknown) of stations in the vicinity and there has to be downlink traffic from the AP. Similarly, in order to keep the overheard information as fresh as possible, there has to be a higher level of medium contention. Practically, none of these assumptions can be guaranteed.

The authors in [6], propose two rate adaptation schemes, both of which are essentially enhancements to previously proposed rate adaptation mechanisms. In their paper, they presented a detailed explanation of various factors which affect the design of a rate-adaptation scheme for WLAN devices, specifically; they discuss the overall system latency and categorize such systems as low latency and high latency systems. System latency has serious performance and design implication on a rate adaptation scheme; for example, per frame rate adaptation can be used only in a low latency system, where the two way communication latency between the PHY (where the transmission status is known) and the rate adaptation scheme (resident at the MAC sublayer and where the information is used for making a decision) is less than 28 µs (which is the DIFS duration for IEEE 802.11g). ARF, which relies on the outcome of the transmission status of a packet, can only be implemented on low latency systems.

Rate-adaptation schemes are widely studied from different perspectives [24]-[30]. An analysis of the impact of various rate adaptation protocols on routing protocols for multi hop wireless networks is presented in [31]. The authors discussed ARF, RBAR and the rate adaptation mechanism as described in the 802.11n draft. In their opinion receiver based rate adaptation mechanisms which they term as closed loop approaches are better than statistical (open loop) approaches. However, in the pre- 802.11n WLANs receiver based systems would require modification to the standard frame formats and therefore were not considered a workable option. The 802.11n draft defines a new control field called HT (high throughput) for this purpose and therefore, according to their findings the rate adaptation mechanism for 802.11n draft performs best when compared to rest of approaches. However, the 802.11n rate adaptation scheme performs poorly in some scenarios because it does not consider MPDU length which is difficult to estimate as a number of MPDUs can be aggregated to form a single A-MPDU.

3.2.2 SNR-based rate-adaptation schemes

Changing the transmission rates essentially change the underlying modulation techniques. For decoding transmissions a receiver requires a certain level of SNR. Therefore, once the existing SNR is known, a suitable modulation technique (and thus a transmission rate) which can be decoded at the existing SNR can be selected [20, 32, 34-42].

SNR-based rate-adaptation schemes uses SNR measurements to estimate the quality of channel and such measurements are utilized for selecting a suitable transmission rate. SNR-based rate-adaptation schemes can provide very accurate channel state information when compared to other schemes where the sender estimates CSI from the delivery-success rates of previously transmitted frames [33].

Ideally, in SNR-based rate-adaptation schemes a transmitter should know about the SNR levels of its transmission at the receiver, however it is not the case. To make it happen various proposed solutions provide runtime feedback about the SNR levels to the transmitter. Rate-adaptation schemes in which a receiver's feedback is used for selecting a transmission rate are called *closed loop systems.* In *closed loop systems* (or more appropriately *receiver-side rate-adaptation schemes*) e.g. the Receiver-based auto-rate (RBAR) [34], the channel quality estimation (through SNR levels) and selection of transmission rate both are performed by the receiver.

3.2.2.1 Receiver-side SNR-based closed loop rate-adaptation systems

Rate-adaptive framing (RAF), [35] is a closed loop rate-adaptation scheme in which a receiver analyzes the SNR and provides a feedback to the transmitter about an appropriate *frame size* and *transmission rate.* In RAF, a receiver piggybacks the optimal transmission rate and frame size in the ACK frame. In order to incorporate the rate and frame size in the ACK frame, RAF, modifies the *duration* field in the ACK frame. In a single frame transfer when no fragmentation is used, the ACK field is set to zero to indicate the end of busy medium condition. In the case of a fragment burst, when second and following fragments of an MSDU are sent right after receiving ACK for previous fragment, the duration field contains the time in microseconds in which the current frame transfer would complete. RAF divides the 16-bit duration field into two subfields, one 4-bit *Channel Rate* subfield and the other 12-bit *Frame Length* subfield; this is shown in Figure 3-2.

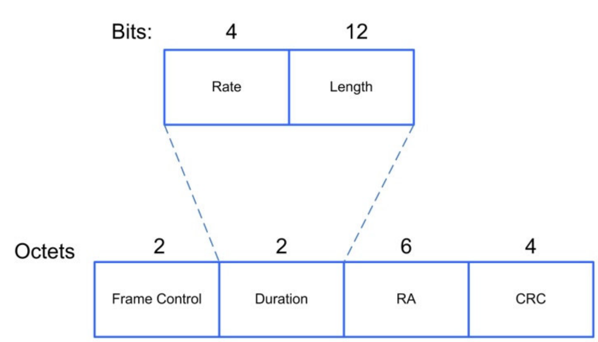

Figure 3-2: RAF's change of the Standard ACK frame to include 'rate' and 'length' feedback to the transmitter.

SNR-guided rate-adaptation (SGRA) [36] uses probe frames in combination with SNR-feedback. SGRA defines two states: *interfered state* and *interference-free state*. While in the interference-free state, the rate-adaptation is solely done on the SNR feedback while in the interfered state the SNR feedback is only used as guide and actual decision is made using probe frames.

Various network applications have different requirements in terms of transmission rate, delay and loss tolerance; this information when communicated to the rate-adaptation module can help it take an intelligent decision [22]. The authors in [22] proposed a scheme which essentially relies on RBAR with a modification that a sender has to specify the *loss tolerance* of the transported traffic in order that the receiver uses both this information and the current channel estimation to select the appropriate transmission mode. The loss-tolerance information is conveyed to the receiver in RTS frames, by using two bits in the standard headers.

Full Auto Rate (FAR) [37], proposed the use of RTS/CTS procedure before every frame exchange sequence. The transmission rate for these frames is selected using ARF like rate-adaptation schemes, where as for the actual data-frames, FAR proposed the use of RBAR. As proposed in this scheme, while sending the RTS frame, the sender is unaware of the transmission rate of the data-frames that it would send and thus does not know the duration for which it would reserve the medium. Therefore, the medium reservation which is done through the RTS frame by the sender is fundamentally flawed.

3.2.2.1.1 Analysis of sender-side SNR-based closed-loop rate-adaptation schemes

Closed-loop rate adaptation schemes provide a better estimate of the CSI and such a rate-adaptation scheme clearly performs better than those schemes which purely rely on 'hit-and-trial' based, sender side schemes. Closed-loop systems are robust and converge quickly to the channel conditions faster than sender side schemes.

However, the persistent problem in all (to the best of our knowledge so far) Closed-loop rate-adaptation schemes is the method through which a receiver conveys its feedback to the transmitter. As mentioned in the above paragraphs, all the existing closed-loop rate-adaptation schemes which rely on feedback from a receiver STA require necessary changes to the original IEEE 802.11 frames specification, e.g. the RBAR requires changes to the MAC data-frame, the RTS/CTS frame and PLCP header. Similarly, RAF changes the duration field in the ACK frames.

Incorporating changes in the standard frames renders such rate-adaptation solutions (e.g. like the RBAR and others relying on run-time feedback from the receiver) incompatible to co-exist with the legacy (standard IEEE 802.11 conformant) STAs in a BSS. This is a serious drawback of closed loop rate-adaptation schemes. To avoid using feedback information from a receiver, SNR levels of ACK frames from a receiver are used to estimate the channel quality at a sender [38-42]. Such rate-adaptation schemes are thus *sender-side SNR-based open loop systems*.

3.2.2.2 Sender-side SNR-based open loop rate-adaptation systems

The initial research which proposed the SNR based rate-adaptation schemes required modification to the standard frames, thus discouraged further research on the better possibilities for SNR-based closed loop rate-adaptation schemes. But nevertheless SNR based CSI is accurate and rate-adaptation schemes relying on SNR are relatively more robust. Therefore, to use the advantages of SNR for determination of CSI and at the same time avoid modification in the standard frames, *sender-side* SNR based schemes were proposed.

A United States patent [38], proposed the use of SNR of ACK frames at the sender side to approximate the SNR at the receiver and select appropriate transmission rate. Inspired from SampleRate and SNR-based approaches, authors in [39] combine frame statistics information with SNR of ACK frames to select a transmission rate which gives highest throughput. A similar approach is used in [40], which uses SNR of beacon frames to estimate the channel conditions and select an appropriate transmission rate thereafter. The solution proposed in [20], records the signal strength of frames received from an access point, and when the average value of the received signal strength passes a certain threshold, the rate-adaptation scheme select the associated transmission rate. The thresholds are adaptively adjusted according to the frame failures at the associated rates.

The authors in [41], proposed the use of frame error rate information for rate adaptation in conjunction with SNR information. Inherently this scheme uses logic similar to SampleRate; where it periodically sends probes channel conditions by sending frames at higher and lower rates. The authors believe that by doing so, without the SNR consideration, can penalize the overall performance in a sense that in case of good channel conditions the frames sent at lower rates cause poor performance and in the case of bad channel conditions, probing with higher data rates can cause frame looses and retransmissions. To cope up with this issue, they use SNR information which guides while probing at lower and higher transmission rates. When the channel conditions are good, as reported by higher SNR, then the channel will be probed with more number of frames sent at higher rate and less with lower rate. Conversely, when SNR is low, a smaller fraction of frames are sent at higher rates thus reducing the number of frame losses and delay.

Similarly, a rate adaptation mechanism using both statistical feedback and RSSI of ACK frames is proposed in [42]. The statistics are continuously collected and decision is made based on the throughput i.e. a transmission rate which provides highest throughput is used with no consideration of the amount of loss it may incur. The RSSI is used to safeguard the selected rate and improve system responsiveness. According to authors ONOE is more conservative and less sensitive to individual packet failures than ARF. In case of good channel conditions it takes approximately 10 seconds to select a higher transmission rate and in case of bad channel conditions it takes it relatively longer to settle down at a non-lossy lower rate and eventually it causes heavy packet losses. It modifies the MADWIFI rate-adaptation scheme by setting the 'c_0' to 2 which means that in case of failures at 'r_0' the retransmission would be repeated at a transmission rate 'r_0'. After the second failure further retries are made at lower rates (r_1,r_2,r_3) with c_1,c_2,c_3 all set to 1. This is to enable the algorithm responsiveness to short-term channel variations. The parameter 'r_0' here is called the long-term transmission rate. The long-term transmission rate for a node is adjusted according to its current state. A station is in either of two states at a time; the "T_x" and "*Probe*" state. At a new round which is determined by certain timers, a station is in the "T_x" state and uses r_0 for transmission. If the channel conditions are good over a period of time the station uses a new (higher) transmission rate and enters into the "Probe" state where the following transmissions are done at *r0=probeRate*. If the transmission at the new rate is a success, the 'r_0' and the long-term transmission rate is set to this new rate and system enters back into the T_x state. After selecting a new value for r_0, corresponding values for r_1,r_2 and r_3 are also adjusted. However, this scheme is essentially like the ARF and the function which determines the rate switching (*may_probe()*) takes a decision based on the successful outcomes of previous frames using a certain threshold. The only difference is that while the transmission rate is less than maximum value and channel conditions are good then it may select a higher rate if the RSSI from ACK frames gives a positive indication. A state transition diagram showing this process is given in figure-3.3.

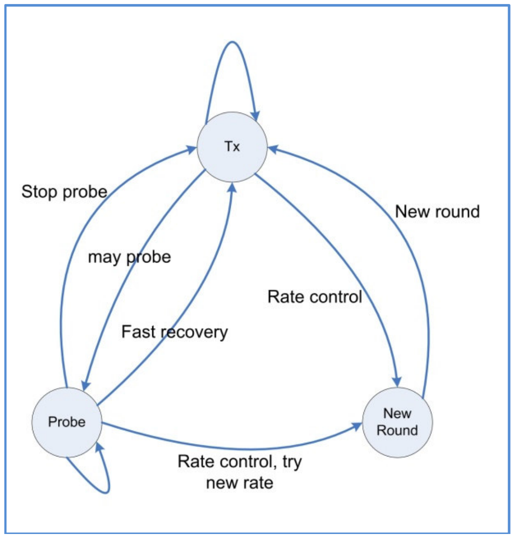

Figure 3-3: State transition diagram used by [42].

3.2.2.2.1 Analysis of sender-side SNR-based open-loop rate-adaptation schemes

Various assumption are involved in designing *sender-side SNR-based open loop systems*, e.g. it is assumed that the SNR levels recorded for the ACK frames (at the sender side) are the equal to the SNR levels that are experienced by data-frames at a receiver. However, this assumption is not always true and can possibly lead to wrong estimation of channel quality due to the following reasons:

1. The transmission rates used for actual data frames are different from that of ACK frames.
2. The size (and thus the transmission duration) of the data frames and the ACK frames are also different.
3. There is a possibility of uneven SNR levels at the two ends because of the presence of hidden terminals or other sources of interference at either of the two sides, which results in wrong assumptions of the SNR levels at the other end of communication.

Therefore, sender-side SNR-based rate-adaptation schemes can avoid the use of feedback delivery by a receiver and thus avoiding changes to the standard frames as done in the previous solutions; however it is done using (a potentially non-realistic) assumption.

3.3 Responsiveness of rate-adaptation

ARF uses a fixed threshold for increasing the transmission rate, owing to which it is unable to react quickly to channel variations. At the same time, if the variations in channel quality are very slow, such fixed threshold values for increasing the transmission rate may force the ARF to overreact by attempting to transmit at higher rates, resulting in retransmissions. To improve over this shortcoming in the ARF scheme, the authors in [43] proposed a dynamic link adaptation scheme which adaptively uses a *short probing* and *long probing* interval to cope up with fast and slow fading wireless channels. The procedure is explained in Figure 3-4. The algorithm uses predetermined thresholds, S, S1 and S2. After S successful transmissions transmission rate is incremented and a transition to an intermediate state '*spread?*' is made. While in the '*spread?*' state, if the next transmission is successful, the algorithm assumes that the channel conditions are improving and transmission at a higher rate is possible, therefore, another transition is made to '*High*' state and S is set to S1 (i.e. S=S1). With the value of S1<S2, it implies that the responsiveness of the rate adaptation algorithm should be high in order to meet the quickly changing (improving) channel conditions. On the other hand if the transmission at the higher rate (when the state is '*probe?*') fails, the algorithm assumes that the previous decision of increasing the transmission rate was premature and as a result a transition to 'Low' state is made and S is set to S2.

Although the two thresholds (i.e. S1 and S2) enables the rate adaptation to improve its responsiveness, nevertheless, the values of these two thresholds are set heuristically which indicates that the responsiveness of the algorithm would be improved for specific variations in the behaviour of the wireless channels but it cannot be generalized for all situations.

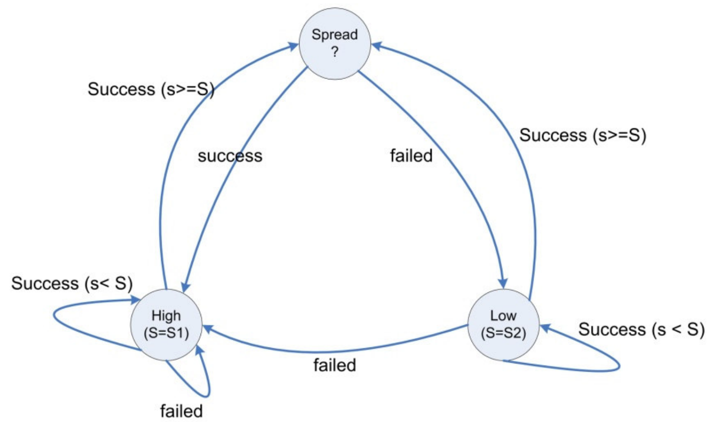

Figure 3-4: State transition diagram as mentioned in [43]. 's' is the total number successful transmissions. 'S' is the success threshold having various values, S1 and S2, where S1 < S2.

The efficiency of a rate-adaptation scheme relies on the mechanism which determines *when* to decrease the transmission rate and *how often* an attempt be made to increase the transmission rate. The authors in [44] aim at creating an adaptive mechanism to direct the transmitter's rate increasing attempts in order to meet a target delay, with minimum frame losses as a result of rate increase. Normally a single frame failure is regarded as failure due to collision but consecutive frame failures are due to deteriorated channel conditions. However, the decision of how often should the transmission rate be incremented is more challenging; because from one perspective when a transmitter increments the transmission rate more frequently it would be able to respond to channel variations, however on the other hand as a result of such tendency there are higher chances to encounter frame losses. The authors in [44] proposes a method which, under given delay factor (response time of the rate adaptation algorithm) calculates the number of consecutive successful transmissions after which the rate adaptation algorithm would attempt to increase the data-rate.

3.4 Types of frame losses

The decisions of most of rate adaptation schemes are based on the transmission status statistics of frames. Frames losses generally depict channel quality deterioration while on the other hand, successful transmission of frames is considered to be a sign of improved channel conditions. Loss of frame (and thus channel quality deterioration) is indicated when a sender STA does not receive ACK for a previously transmitted frame.

However, when carefully examined, such losses can occur because of two different reasons and each requires a different action from a rate adaptation scheme.

3.4.1 MAC level losses (Frames losses because of collisions)

Collisions occur when two or more than two STAs simultaneously transmit frames so that an intended receiver is unable to decode the transmission and thus unable to ACK back to the sender. Simultaneous transmissions can occur because of (i) failure of the MAC protocol in highly congested BSS or (ii) because of operation of hidden nodes in the vicinity of a transmitter. Such losses are indicative of the fact that the (PHY) channel conditions are supportive and unchanged and the transmission is corrupted due to a failure at the MAC layer; therefore, such losses of frames are called MAC level frame losses [45].

a. Collisions in highly congested networks

In a highly congested BSS the percentage of frame losses because of simultaneous transmission of stations is reported to be as high as 30% [46]. The scenario reported in [46] used network-traffic's traces from the network which was setup for the 67th Internet Engineering Task Force (IETF) meeting, in 2006. In this scenario, if ARF, AARF, SampleRate as reported in [16] or any other rate-adaptation scheme which relies on frame success statistics would decrease the transmission rate. For instance, 30% loss implies that 3 frames out of every 10 frames have to be retransmitted. ARF and AARF increase the transmission rate only when the number of successfully transmitted frames reaches 10 or more (in the case of AARF), therefore, in this scenario, there is no possibility of increasing the transmission rate if ARF/AARF is used in the client stations. On the other hand there are higher chances that out of 3 frames failures in every 10 frames there would have been 2 consecutive frame failures. ARF and AARF both, lower the transmission rate when two consecutive frame losses occur. Therefore, it is likely that if the frame-loss rate as reported in this scenario persists, ARF like rate-adaptation schemes would ultimately reduce the transmission rate to the lowest. Similar observations are reported in [16], where the authors recorded that 73% of the total frames were transmitted at the lowest rate. Similar studies such as

[47], [48], and [49] have reported that most of the rate-adaptation schemes do not perform well in highly congested wireless networks. Rate adaptation schemes which do not use a frame-loss differentiation mechanism wrongly associate such frame losses to channel quality degradation. Lowering the transmission rate, at times of high medium congestion causes further performance degradation.

b. Collisions because of hidden nodes

Frame losses can also occur because of simultaneous transmissions of frames as a result of presence of hidden stations. The phenomenon of hidden stations arise when a receiver in the middle of two transmitter receives two frames from both transmitter and is unable to receive either, thus drop the frame and none is acknowledged back to the sender. The senders in this scenario are out of the each other's coverage areas and thus their Clear Channel Assessment (CCA) mechanism cannot detect busy medium condition even when one of the two senders has acquired the medium.

In case of such losses, reducing the transmission rate further deteriorates the communication efficiency.

3.4.2 PHY level losses

a. Corrupted frames

During frame transmission, a frame can be corrupted due to reduction in signal quality e.g. lower SNR, as received by a receiver. The Cyclic Redundancy Check (CRC) field in the PPDU, in this case indicates that the frame contents are wrongly received and not as sent by the sender. Such frames are discarded with no ACK or feedback to the original sender.

b. Totally lost, undetected frames

In this case, the energy level of the received signal at the receiver is lower than the receiving threshold of a receiver, rendering a receiver unable to detect transmission and decode contents of a PPDU. Such transmissions are completely lost and the receiver has no knowledge and thus cannot provide any feedback to the transmitter.

It is important to highlight that during a frame exchange sequence, a data-frame and an ACK-frame are exchanged between a sender and receiver and either of these two types of frames can be corrupted. In other words, a frame loss can be because of: a collided data-frame, collided ACK-frame, corrupted data-frame, corrupted ACK-frame, totally lost data-frame or totally lost-ACK frame. Using the same rules which formed the fundamentals of most of the rate adaptation schemes and which can be stated as '*transmission at lower rates makes frame more robust to corruption*'; we can assume that out of the six likely reasons (as mentioned above) for frame losses (and thus indicating poor channel conditions), the corrupted-ACK frame is least likely to occur; because ACK frames are always sent at comparatively lower transmission rates. However, the other five are equally likely to occur.

MAC level frame losses need to be differentiated from PHY-level frame losses as they require a different (more specifically no action) from the rate-adaptation scheme. In case of no loss-differentiation, a rate-adaptation scheme would generally associate every frame loss with deteriorated channel conditions, which in many cases would require the rate-adaptation scheme to lower the transmission rate. If the frame losses are because of simultaneous transmissions (i.e. MAC level losses) then lowering the transmission rate would further degrade the overall performance. Therefore, it is important to devise a mechanism to differentiate frame losses before a rate-adaptation mechanism takes action.

3.5 Loss-differentiation

A number of loss-differentiation mechanisms are proposed in the literature, most of which deals with differentiation of MAC level losses from PHY level losses [50-61]. The authors in [50] evaluated the effectiveness of loss-differentiation mechanisms in various scenarios with varying network loads. An analysis of the effect of contention on the performance of rate-adaptation schemes is presented in [51]. The transmission characteristics are affected by a number of factors, each of which has to be considered while devising a rate-adaptation strategy [62]. According to the method proposed by the authors in [52], a *consecutive failure count* '*n*' is compared with *probe activation threshold*

'P_{th}' and *consecutive failure threshold 'Nth'*. When *n* reaches P_{th} , the next data frames will be sent with RTS/CTS frames and when *n* reaches N_{th} the next data frames will be sent at a lower rate. The default values for P_{th} and N_{th} are 1 and 2 respectively. This scheme uses a threshold of 10 for successful transmissions and after that the next transmission is done at a higher rate. Using those values for the three different thresholds; this scheme is essentially like the ARF with only one change that after the first frame loss, it will use RTS/CTS and after the second frame loss it will reduce the transmission-rate.

The authors in [52] also proposed the use of CCA to identify the reason of frame losses. This is done in a scenario when a STA sends a data frame and waits for ACK. In normal condition the CCA should not indicate *busy channel* condition, however while a STA is waiting, and the CCA indicates busy channel it is concluded that a collision has occurred. In this case, the STA would retransmit without increasing the failure count or lowering the transmission-rate. This mechanism would not work if there is a simultaneous transmission from other stations and their transmission ends before or after the transmission of a station under consideration. In both of these cases there would be frame losses because of collision which can be detected with the help of RTS/CTS mechanism.

An RTS threshold is used in [45] to activate RTS/CTS handshake when the duration of a data-frame crosses the RTS-threshold. A similar RTS/CTS technique is used in [55], with ARF. To improve over the trial based or threshold-based usage of RTS/CTS, authors in [56] proposed probabilistic approach for activating the RTS/CTS procedure for loss differentiation. An RTS-probability parameter is maintained, whose value is calculated by using the collision probability of transmitted frames using a mathematical model.

The use of RTS/CTS causes an overhead especially when the MSDU size is small and this overhead can affect the performance while transmitting Voice over IP (or other real time traffic) [63],[64]. As an alternative to using RTS/CTS for loss-differentiating, the authors in [40], suggest the use of lowest transmission rate right after a frame loss. According to [40], if the receiver is still within the range, then transmission at lowest rate would make it possible that the receiver would receive the transmission and send back an ACK. This condition would indicate that transmission failure was because of channel quality deterioration. On the other hand, if the transmission at the lowest rate is lost (not acknowledged) however, the sender can still receive beacon frames; then this condition would indicate that frame losses are due to collisions.

3.5.1 Analysis of frame-loss differentiation mechanisms

The outcome of a loss-differentiation mechanism tells us that whether the frame losses are because of collisions or due to channel quality variations. As a general trend in most of the related research literature, it is assumed that collisions are only because of hidden-nodes and as a result RTS/CTS frames are used before actual data-frame exchange. However, as reported in [46], the percentage of frame collisions because of simultaneous transmission increases as the station density increases in a BSS. In such cases, even if the loss-differentiation mechanism identifies the actual reason of frame loss as to be because of *collisions*, it doesn't always imply that the collisions are because of hidden-nodes. Therefore, after loss-differentiation, simply using RTS/CTS for collision avoidance is not the correct strategy and would cause extra overhead. This problem was investigated by [46], where the authors tried to find a correlation between frame losses and *channel busy time* (CBT). CBT was used as a measure of medium congestion. Although, in general the number of frame losses increases with increase in CBT, but there no clear correlation between the two. Therefore, a mechanism which can clearly identify the difference between the two kinds of collision losses can tremendously improve the decisions of a rate-adaptation scheme.

Lastly, to the best of our knowledge almost all frame-loss differentiation mechanisms rely on the use of RTS/CTS procedure. The use of RTS/CTS itself is an overhead and quite unfortunately it requires several frame exchanges (with and without RTS/CTS) between the sender and a receiver to arrive at the final conclusion. Therefore, a loss-differentiation procedure which minimizes the overhead of RTS/CTS and at the same time improve its own responsiveness would greatly improve the overall communication efficiency.

3.6 Loss-differentiation and MAC Backoff counters

The IEEE 802.11 standard uses randomly selected backoff, which is equal to the number of randomly selected timeslots that a station has to wait when the medium is idle. This is to avoid the possibility of simultaneous access to the medium by two or more stations causing collision of frames. When a frame is not acknowledged the standard assumes that the frame loss occurred because of collision and as a result of that it doubles the interval which is used for selecting a random value of the backoff.

Frame loss differentiation mechanisms can identify the actual reason of frame losses. It is logical not to increment the backoff values if the frame loss is because of channel quality degradation [40]. By introducing such an enhancement the performance of MAC can be improved as the contention delay is minimized.

3.7 Effect of Retransmissions

Avoiding the number of retransmission is very important. In order to analyze the overhead that is rendered to the communication as a result of transmission failure, it is worthwhile to analyze the procedure followed by the MAC protocol during a retransmission. Before the transmission of every frame a station waits for the medium to be idle for DIFS interval, after which it waits for the random backoff counter to reach zero and if there is no other station using the medium, it starts transmission. After transmitting a frame a station expects to receive ACK within a predefined interval of time, called *ACK-Timeout* which is equal to the sum of SIFS time, SlotTime and PHY-RX-START-Delay (as given in equation 3.1)
If the sender does not receive an ACK frame it assumes that the transmission failed, therefore, it increments the appropriate retry count (either of the *short retry* or *long retry count*) and at the same time the station increments the contention-window (CW) parameter to next higher (twice its previous) value. So, that the random backoff which is selected before the next transmission may possibly have a higher backoff value, which implies that a station would wait for longer time before acquiring access to the medium.

Therefore, as a result of higher retransmissions, a communication process suffers from longer delays, causes higher medium occupancy (poor utilization of the shared spectrum) and consumes higher energy while transmitting a unit of data.

3.8 Rate-adaptation and energy efficiency

A rate-adaptation scheme can play an important role in determining the energy efficiency of the communication process [65-66]. Ideally transmission at the highest possible rate, with minimum number of retransmissions is extremely efficient in terms of per bit energy consumption. However, this is not always possible as transmission rates are dynamically adjusted and operation at each rate causes a variable number of retransmissions.

Retransmission of a frame means that medium has to be acquired again to repeat the same process, thus causing spectral inefficiency. In most of the rate-adaptation schemes (which determine the CSI through the transmission success/failure of frames), retransmissions cause reduction of transmission rate. Therefore, after retransmission the same frame is transmitted at a lower rate, which means that the transmitter has to operate for a longer duration. Over a certain period of communication, retransmissions' cumulative effect is a higher inefficient use of battery power.

Therefore, rate-adaptation schemes which reduces (and to a larger degree avoid) retransmissions can effectively enhance the energy efficiency of a communication system.

3.9 MAC timing constraints and effects on rate-adaptation schemes

From functional perspective rate adaptation algorithms are functionally a part of the MAC sub-layer. The implementation of MAC sub-layer functionality is device specific and involves combination of dedicated hardware and software. The design and combination of hardware and software for MAC implementation influences the communication latency and thus it also affects the rate-adaptation algorithm.

The authors in [6] identified two classes of 802.11 standard devices: *low-latency* and *high-latency systems*. In low-latency systems per-frame rate adaptation algorithms can be implemented while high-latency systems require periodic analysis of channel state and update certain parameters which enable rate adaptation. For low-latency systems to work the communication latency between the Physical layer and the rate control algorithm should be less than $T_{success}$. According to the authors $T_{success}$ is the time interval between the end of a successful transmission and start of another and is equal to the distributed inter-frame space (DIFS) the smallest value being 28 µs for IEEE 802.11g.

However, we cannot term it a per-frame rate control algorithm if we consider DIFS as the time interval for decision making in a rate controller. The reason is that in a fragment bust (please refer to chapter-2), various fragments of an MSDU are sent back-to-back and the ACK are spaced by an SIFS interval. Therefore, for a rate-adaptation scheme to be called a per-frame rate-adaptation scheme it should be able to react within the SIFS interval of time. The rate-adaptation schemes which uses a DIFS interval as a reaction time can rightly be called as per-transmission rate control mechanism. The timing detail for different versions of IEEE 802.11 standard is given in Table 3-1.

Table 3-1: Timing constraints for design of a rate-adaptation scheme.

PHY	SIFS duration in µs	DIFS interval in µs
FHSS	28	128
DSS	10	50
IR	7	11
OFDM	16	34

3.10 Analysis

Frame-statistics based rate-adaptation schemes are unable to determine the duration of channel variation which is very important in formulating their strategy of rate-adaptation. To elaborate on the importance of duration of channel variation and its effect on the strategy of a rate-adaptation, consider performance of SampleRate in short term channel variations. Assume that the transmitter was operating at 11 Mbps when the channel condition started to deteriorate and 11 Mbps is the most suitable transmission rate between the transmitter and receiver pair in best conditions.

At the point when the short term variation starts and transmission at 11 Mbps is no longer possible, SampleRate would lower the transmission rate after 4 successive failures, e.g. to 5.5Mbps. After 10 successful frame transmissions, SampleRate would try a higher transmission-rate value, which in this case is higher than 11 Mbps. The reason for trying a higher transmission rate (than 11 Mbps) is that SampleRate doesn't use a rate which caused transmission failures recently. Thus 11Mbps just failed and it is marked as a value which will not be sampled for some time. If the channel recovers from errors at this point (as it was a short term variation), so that the transmitter can again successfully transmit at 11Mbps. However, SampleRate, would attempt to use an even higher transmission rate, which in the currently assumed scenario would fail. And as result in this scenario, SampleRate would decrease the transmission-rate back to 5.5 Mbps after 4 frame failures at the higher rate. This process would continue and after every 10 successful transmissions (at 5.5Mbps), it would attempt to select a higher value and cause 4 frame losses, till the time when the 11Mbps data-rate value is marked 'clean'.

In such scenarios, ARF and AARF would perform better than SampleRate because after the first four failures when the transmission rate is lowered to 5.5 Mbps, SampleRate would send 10 frames at 5.5Mbps, 4 at 24 (or higher) Mbps which would be retransmitted again at 5.5 Mbps along with 6 more frames at 5.5 Mbps, after which it would try it at 11 Mbps (when the 11 Mbps transmission rate is marked 'clean' again). So, a total of 20 frames would be sent at 5.5Mbps along with 4 transmission failures at 24Mbps and 4 failures at 11 Mbps. While in the same situation ARF and AARF would send 10 frames at 5.5Mbps and 10 at 11Mbps with 2 retransmissions only in the beginning when the short term channel variations started. Fig-1 shows SampleRate and ARF in short-term channel variation scenario.

Similarly, consider a long term channel variation situation; assume that the transmitter is again operating at 11 Mbps when the channel conditions started to deteriorate. In this situation SampleRate would lower the transmission-rate to 5.5 Mbps and after 10 frames, it would try 24 Mbps or higher which would fail again and it would operate back on 5.5 Mbps. This process would continue and SampleRate would randomly select higher transmission-rates (alternatively excluding some values at which recent transmissions failed). In this situation, AARF would perform better because after each failure it would stay for a longer time at the 5.5 Mbps before attempting to operate at higher data-rate. ARF, on the other hand would perform similar to SampleRate, in terms of the frequency of rate increasing attempts. When the conditions get back to normal it may take SampleRate and AARF relatively longer time to operate at 11Mbps while ARF will soon resume operation at 11Mbps. While in the same situation ARF and AARF would send 10 frames at 5.5Mbps and 10 at 11Mbps with 2 retransmissions only in the beginning when the short term channel variations started. Figure 3-5 shows SampleRate and ARF in short-term channel variation scenario.

Similarly, consider a long term channel variation situation; assume that the transmitter is again operating at 11 Mbps when the channel conditions started to deteriorate. In this situation SampleRate would lower the transmission-rate to 5.5 Mbps and after 10 frames, it would try 24 Mbps or higher which would fail again and it would operate back on 5.5 Mbps. This process would continue and SampleRate would randomly select higher transmission-rates (alternatively excluding some values at which recent transmissions failed). In this situation, AARF would perform better because after each failure it would stay for a longer time at the 5.5 Mbps before attempting to operate at higher data-rate. ARF, on the other hand would perform similar to SampleRate, in terms of the frequency of rate increasing attempts. When the conditions get back to normal it may take SampleRate and AARF relatively longer time to operate at 11Mbps while ARF will soon resume operation at 11Mbps.

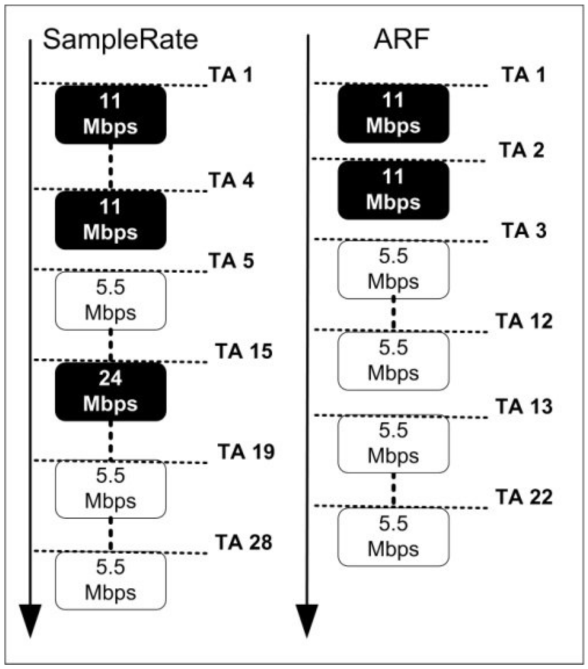

Figure 3-5: Comparison of SampleRate and ARF in a specific short-term channel quality variation scenario.

3.11 Important factors in design of rate-adaptation schemes

The performance of wireless networks is highly variable and depends on a number of factors e.g. mobility, distance between the transmitter and receiver, presence of hidden nodes, number of contending stations within the same BSS, interference from neighbouring 802.11 networks and so on. Likewise, there is a rich variety of network applications; having dissimilar requirements from the underlying networking protocols in terms of delay, throughput and mobility. The prevailing tendency of achieving higher throughput by increasing the transmission-rate among majority of the existing rate adaptation algorithms might sometimes achieve positive results. However, it is normally done at the expense of relatively higher delays as transmissions may frequently fail at higher rates. While some applications may be favoured by higher throughput, others will be penalized by the resultant higher delay. Therefore, apart from controlling the transmission-rate according to the changing channel conditions the rate adaptation mechanism should be capable of fulfilling requirements of various classes of network applications.

In order to handle, a similar scenario [67] proposes a method in which three rate adaptation algorithms are used. The rate adaptation algorithm is integrated in a cross layer framework, where a wireless adaptation layer (WAL) communicates important system parameters e.g. type of traffic, rate of various types of traffic, application preferences and also MAC layer indication. WAL adaptation layer selects one of the three algorithms according to the type of traffic which suite the traffic requirements while dynamically selecting data-rate. The solution in [67] is lucid and very logical but the rate adaptation algorithm designed for real-time applications do not take into account the timing constraints of the MAC protocol. Without this consideration (e.g. the backoff and contention delays, etc. at the MAC layer) it will be very uncertain when the algorithm tries to over match the traffic requirements and thus the requirements will not be fulfilled. To support this argument consider Figure 3-6 showing five scenarios with varying number of wireless nodes. In all five scenarios the transmission-rate of transmitter remained fixed. Because of the higher contention delays and higher backoff the amount of data transmitted by a source to a particular destination decreases as the number of participating stations increase. Therefore, without taking care of the timing constraints and limitations of the underlying protocols, merely overmatching the data-rate requirements (as estimated at the transport layer) will not be useful.

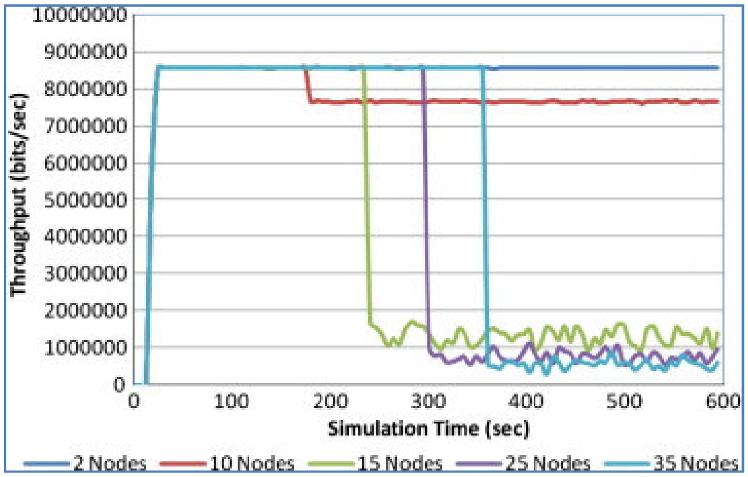

Figure 3-6: Throughput comparison of various scenarios with varying device density.

3.12 Cross-layer rate adaptation (CLRA)

3.12.1 Motivation

A rate adaptation mechanism is required for enabling the IEEE 802.11 MAC to adaptively select a transmission-rate according to varying channel conditions such that the selected transmission rate yields highest achievable performance in that situation. Nevertheless, the reality is that there are a number of factors due to which the performance of wireless networks can be affected. Likewise, various network applications have different requirements which require adaptive tweaking of network performance metrics.

Therefore, it is important that a rate adaptation mechanism should be capable to do the following:

1. Quickly react to the variations in wireless channel's performance.
2. Identify the actual reasons of variations and act accordingly.
3. Select an efficient adaptation interval (which determines how often the rate adaptation mechanism will probe/sense the channel's performance in order to recover after frame losses or at times of continuous good channel conditions).
4. Perform rate adaptation according to application preferences.

Section 3.12.2 gives details of our proposed solution which is essentially a statistics based rate-adaptation mechanism where the core idea is based on *on-demand incremental* *strategy*. The on-demand incremental strategy is motivated by the research and analysis of the current rate-adaptation solutions and their common drawbacks. A very common trait in the design of all rate-adaptation schemes is that they have an inherent tendency of operating at the highest transmission rate. At this point, our analysis shows that it is not always a requirement (the term 'requirement' carries many meaning and depends on the context of layer's where it is referred; it will be elaborated in the following paragraph) to operate at the highest transmission rate. And as shown in the previous section, this tendency is one of the major reasons for causing higher retransmissions. A method to identify the prevailing requirements and the underlying communication constraints would help define an instantaneous limiting value which can be used by the rate-adaptation scheme as the highest transmission rate. By doing so, we can essentially limit the rate-adaptation scheme to avoid unnecessary attempts of transmission at higher rates and thus avoid the potential retransmissions which are highly likely to occur at higher transmission rates.

According to the on-demand incremental strategy, a station should only select transmission rates which can meet the outbound traffic demands of its network applications. Therefore, a station, where the applications generate an outbound traffic at the rate of e.g. 2 Mbps; the rate-adaptation scheme should not attempt to operate at the highest (54 Mbps) transmission rate. This would reduce the number of retransmissions which are highly likely especially in the range of 24-54 Mbps (the frame loss rate, in this case depends on the mobility patterns of the source and destination stations, the number of contending stations and other sources of interference in the close proximity of the communication stations). However, there are two conditions which can arise as a result of the on-demand incremental strategy:

- For a station, whose outbound traffic requirements are very low (i.e. less than 1 Mbps) would then use the lowest transmission rate, which means that this station would use the medium for a longer duration of time and would thus penalize other stations which will also be contending for the medium. Therefore, the on-demand incremental strategy should avoid this condition and should have a mechanism of incorporating the medium contention information while calculating the limiting value for the transmission rate.

- A station whose outbound traffic requirements are e.g. 1 Mbps, and the device density the proximity of that station is higher, then the limiting value of the transmission rate which can enable the station to meeting the outbound traffic requirements would be way higher than 1Mbps; it is more or less directly proportional to the medium occupancy.

In order to cope with the above mentioned conditions which result from such a design, we formulated a mechanism to calculate at run-time, the higher layers traffic requirements and the underlying MAC sublayer's constraints. A limiting value for the transmission rate is then calculated using such constraints and requirements, and is passed on to the statistical rate-adaptation scheme. The statistical rate-adaptation scheme use this dynamic limiting value for selecting a suitable transmission rate.

The rate adaptation mechanism is integrated in a cross layer framework. The channel conditions are constantly monitored using a statistical approach (using the number of frame losses as a measure to identify the channel state). However, loss differentiation is used in order to make intelligent decisions about assessing channel quality. Consequently, variations in channel performance because of user movement, changes in the environment or presence of hidden nodes in a BSS can be differentiated before rate adaptation. The proposed rate adaptation closely matches applications requirements while keeping in consideration the timing constraints of the underlying network protocols.

3.12.2 CLRA rate-matching

The CLRA is essentially based on a rate matching (matching outbound traffic rate with effective transmission rate) approach. Therefore, to establish an analytical model for analysis of CLRA, it is important to consider the MAC sub-layer timing constraints.

For ease of explanation, time is divided in blocks where a time block includes the contention delay, DIFS interval, backoff duration and time for data transmission, as shown in Figure 3-7.

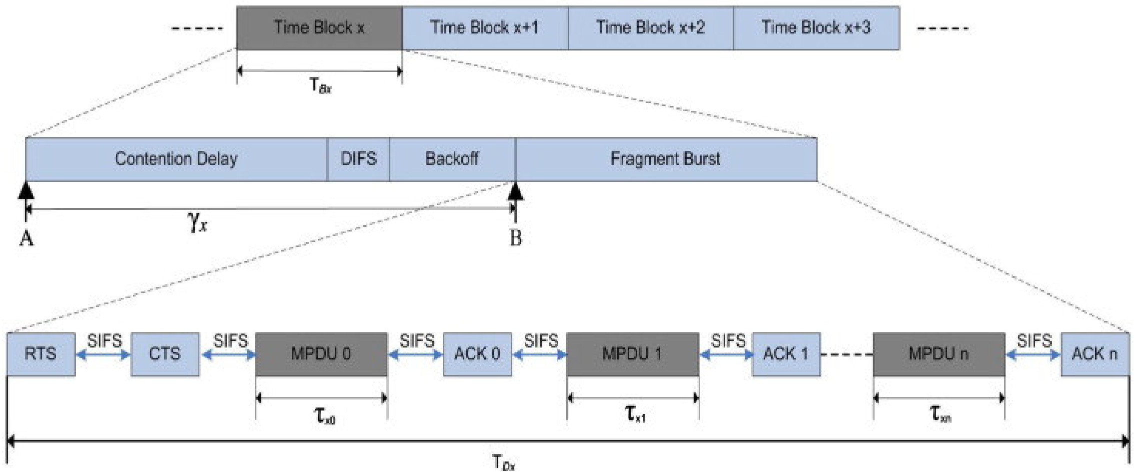

Figure 3-7: IEEE 802.11 MAC-timing block diagram.

Consider a time block ‘x’ of duration T_{Bx}, where the contention delay is T_{Cx}, backoff duration is β_x and the time required for transmission of data during T_{Bx} is T_{Dx}.

Therefore,

$$T_{Bx}=T_{Cx}+mDIFS_x+\beta_x+T_{Dx} \tag{3.2}$$

Here, duration of every T_B may vary because of the variable T_C, single/several DIFS intervals (denoted by m, where m={1, 2, …CW}) and β.

Let the transmission time of an ‘*i*’ th MPDU during T_{Bx} is ‘t_{xi}’ where *i = {0, 1, …,n},* and ‘n’ is the total number of fragments (MPDUs).

The value of ‘t_{xi}’ depends upon the MPDU size and the transmission rate used by a transmitter. According to the IEEE 802.11 standard, all MPDUs except the last one of a fragmented MAC Segment Data Unit (MSDU) are of equal size and therefore ‘t_{x0}’ to ‘$t_{x(n-1)}$’ will be equal (given that the transmission-rate remains the same). Therefore, the total effective time for transmission of data within a particular ‘T_{Bx}’ can be denoted by ‘T_{Ex}’ where

$$T_{Ex}=t_{x0}+t_{x1}+t_{x2}+\ldots t_{xn} \tag{3.3}$$

The maximum size of data which can be transmitted within a certain ‘T_{Bx}’ is equal to the size of an MSDU (in case of no fragmentation) or sum of size of all MPDUs (in case of fragmentation). Let us denote the size of transmitted data and the corresponding protocol overhead during a 'T_{Bx}' by 'δ_x'. Therefore, the operational transmission-rate of a transmitter denoted by 'η_T' is

$$\eta_T = \frac{\delta x}{T_{Ex}} \tag{3.4}$$

Although the transmitter may be operating at a higher transmission-rate but because of the timing overhead (contention, backoff and inter-frame space delays) imposed by the MAC protocol the effective data-rate denoted by 'η_E' is essentially lower than η_T because η_E is equal the rate where the same amount of data (δ_x) is transmitted within a longer period of time:

$$\eta_E = \frac{\delta x}{T_{Bx}} \tag{3.5}$$

As, the value of $T_{Bx} > T_{Ex}$, therefore, $\eta_E < \eta_T$.

If the outbound traffic (generated by the higher layers and received by the MAC sublayer for transmission) is generated at a rate denoted by η_{HL}.

Then within a T_{Bx} duration the amount of data arriving from higher layers is = $(\eta_{HL} \times T_{Bx})$.

The rate-adaptation scheme is responsible to adaptively select η_T in such a way so that the amount of data generated by higher layers during any 'T_B' can be transmitted with in that period of time.

$$\eta_T \times T_{Ex} = \eta_{HL} \times T_{Bx} \qquad \text{or} \qquad \delta_x = \eta_{HL} \times T_{Bx} \tag{3.6}$$

This relation can only hold when the amount of data arriving from higher layers is less than or equal to the MSDU size within any T_B. Considering Figure 3-7, let

$$T_{Cx} + \beta_x + m_x \text{ DIFS} = \gamma_x \tag{3.7}$$

Then,

$$T_{Bx} = \gamma_x + T_{Dx} \tag{3.8}$$

For transmitting n-MPDUs, using RTS/CTS mechanism, the timing overhead imposed by the MAC sublayer is equal to the sum of (1 + 2n) times SIFS, 'n' times ACKs, 1 CTS and 1 RTS duration, i.e.:

$$T_{Dx} = T_{Ex} + (1+2n) \times \text{SIFS} + n \times \text{ACK} + \text{CTS} + \text{RTS} \tag{3.9}$$

Let, $$\xi_x = (1 + 2n) \times \text{SIFS} + n \times \text{ACK} + \text{CTS} + \text{RTS} \tag{3.10}$$

If RTS/CTS is not used then $$\xi_x = (2n - 1) \times \text{SIFS} + n \times \text{ACK} \tag{3.11}$$

Therefore, (3.9) becomes:

$$T_{Dx}=T_{Ex}+\xi_x \tag{3.12}$$

Replacing 'T_{Ex}' in 3.12 with $\frac{\delta x}{\eta_T}$ and then using value of 'T_{Dx}' from 3.12 in 3.8:

$$T_{Bx}=\gamma_x+\frac{\delta x}{\eta_T}+\xi_x \tag{3.13}$$

Using 'T_{Bx}' from equation-3.13 in the second form of equation-3.6:

$$\eta_T=\frac{\eta_{HL}\delta_x}{\delta_x-\eta_{HL}(\gamma_x+\varepsilon_x)} \tag{3.14}$$

The rate-adaptation algorithm has to select from a discrete set of values allowed under a particular 802.11 PHY (e.g. 1, 2, 5.5, 11 upto 54 Mbps). Therefore, values obtained from equation-3.14 are rounded up to the nearest value in the transmission-rate set.

Here, 'ξ' is dependent on the number of MPDUs during a fragment burst. As a standard rule of operation in 802.11 based networks, every transmitting station indicates the duration of usage of wireless medium in its outgoing MAC frames. In order to allow the transmitting station to calculate the duration of usage of the medium, the Standard mandates that the responding station shall transmit its control response frames (CTS/ACK) at the same rate as the immediately previous frame in the frame exchange sequence or else at the highest possible rate belonging to the PHY rates indicated by the BSSBasicRateSet element. The transmitting station will have the information about the number of anticipated ACKs in response to the number of transmitted frames and their corresponding inter-frames spaces; and assuming that the responding station will use the highest rate for transmission of CTS/ACK frames, the duration of 'ξ' can be calculated for a particular T_B.

'δ_x' is dependent on the MSDU size and on the fragmentation threshold.

3.12.3 γ-Estimation

The parameter 'γ' is variable and may have different values in every 'T_B'. The duration of 'γ' provides important information about the current channel and BSS characteristics. In general it depends on:

- 'T_C', which in turn depends on the number of contending stations and channel conditions.
- DIFS, it has a fixed value per PHY.

- 'β' depends on the status of previous transmissions which in turn depends on channel conditions and on the number of contending stations.

In order to measure 'γ' within a certain 'T_{Bx}', a timer is started whenever a station has a frame for transmission, or at the end of successful reception of ACK for the last MSDU during a previous time block ($T_{B(x-1)}$) when the station has many frames for transmission. During T_{Bx} the timer is stopped when the backoff counter reaches a value of zero and station can start transmission, as shown in Figure 3-7. The interval between 'A' (starting point of timer) and 'B' (when the timer is stopped) is the total duration of 'γ_x' during 'T_{Bx}'. Higher values of γ may be because of higher device density inside a BSS or transmission failures which may cause higher backoff counters. Every time when medium is found busy, the backoff counters are frozen within all contending wireless stations. The worst scenario where 'γ' has a larger value is when a wireless station has to freeze its backoff counter several times until it reaches zero. In such a scenario, there may be several DIFS and longer contention delays (when other stations are using the medium) and as a result the overall TB will be considerably longer. Consequently, the CLRA would have to select a considerably higher value of 'η_T' even for a small value of 'η_{HL}'.

3.12.4 Constraints

It is important to emphasize that the relationship given in equation-3.14 which is based on equation-3.6 can only hold when the amount of data arriving from higher layers is less than or equal to the 'δ'. However, if 'γ' is considerably longer and/or 'η_{HL}' is close to the highest values of 'η_T' (various PHY support different values for η_T), then the amount of data coming from higher layers may be more than the amount of data which could be successfully transmitted within a particular 'T_B'.

In other words, such a scenario implies that the MAC sublayer cannot match the higher layer's traffic requirements within the physical limits of an 802.11 wireless station. In such situation, the algorithm would keep 'η_T' at the highest transmission-rate value and the application should be triggered to lower the load through a MAC-application cross-layer interface.

3.12.5 Estimation dilemma

In order to evaluate equation-3.14 during a 'T_{Bx}', it is critical to provide a correct value of 'η_{HL}'. However, the required value of 'η_{HL}' during a 'T_{Bx}' has to be determined based on the exact value of 'T_{Bx}'. But 'T_{Bx}' can only be determined if equation-3.14 is evaluated and 'η_T' is determined. Therefore, to solve this predicament, the average value of 'η_{HL}' during previous 10 time blocks are used as an estimate of 'η_{HL}' during the current time block, i.e.

$$\eta_{HL_x} = \frac{\sum_{i=x-10}^{x-1} \eta_{HL_i}}{10} \tag{3.15}$$

Here, we use average of 10 previous values; increasing this value further will disable the ability of this algorithm to respond to sudden rise or fall in data-rate of higher layers. Similarly, we believe that decreasing it further will not truly reflect the current estimation.

3.12.6 Cross-layer information exchange

There is a variety of network applications; every application has different requirements in terms of delay, throughput and resilience to packet losses etc. Therefore, apart from matching higher layer's traffic rate it is important to convey application preferences (delay thresholds, packet losses, etc.) to the rate adaptation algorithm. An argument can be raised as how to convey individual application's preferences to the MAC sublayer and even if it is done, it may further complicate the overall system design. However, a logical solution to this concern is that applications are categorized in to a limited and discrete number of categories as it is done in the case of IEEE 802.11e. Each application category has its own parameter (various performance parameters e.g. delay, internal priority, etc.) configuration. Every network application belongs to any one of the categories. Therefore, instead of conveying individual application preferences to the MAC sub-layer, information relating to a category of applications are conveyed.

Similarly, it is important to inform the applications about the current status of MAC sub-layer in the form of MAC-specific parameters e.g. the highest transmission-rate which can be supported, number of packet losses and average delay, etc.

Such two way information can be conveyed with the help of cross layer interfaces which extend across the protocol suite. A number of enhancements and performance

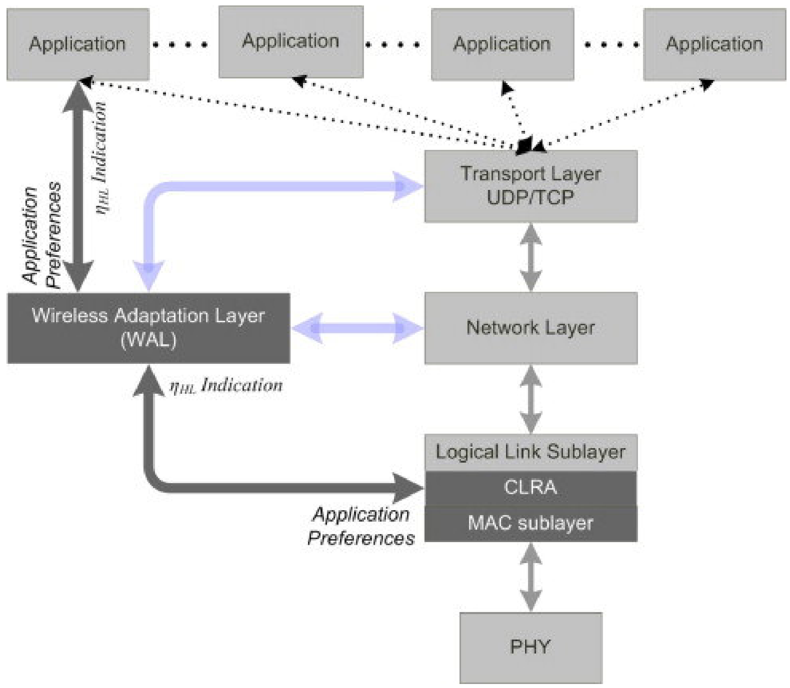

Figure 3-8: Cross-layer communication framework.

improvements which can be introduced with the help of cross layer communication are explored in [68-72].

An overview of the proposed cross-layer communication framework is given in Figure 3-8. Here, a Wireless Adaptation Layer (WAL) is responsible for the cross layer communication among various layers. The functionality of WAL can be embedded in various modules across the protocol suite both in simulation modelling softwares and practical implementation. However, we represent it as a separate entity or a central database of various network parameters with the aim to highlight its role and to keep it in conjunction with our previous work. As mentioned previously, all successfully received frames are acknowledged by the receiver in 802.11 based networks. The time interval between the arrivals of a frame from higher to the MAC layer till the time when an ACK is received provides an easy way of estimating the delay experienced by frames. The actual delay estimation used herein is then calculated by subtracting the time of reception of an ACK frame and two SIFS durations from the time interval mentioned above. Apart from conveying network parameters (e.g. delay), the rate adaptation algorithm also provides an indication to the WAL about the optimum value of 'η_{HL}' which can be supported by the underlying protocols. The indicated value of 'η_{HL}' can be derived from:

$$\eta_{HL} = \frac{\eta_T \delta_x}{\delta_x - \eta_T(\gamma_x + \varepsilon_x)} \tag{3.16}$$

Here, 'η_T' has either the highest value of transmission-rate supported by a PHY, or has the value which result in a higher delay more than the threshold value. In the former case, the current value of 'η_{HL}' is such that it can no longer be supported by the underlying protocols and the relationship given in equation-3.14 is void. Therefore, equation-3.16 provides a recommended value of 'η_{HL}' which can be supported by the MAC layer at a given time. The latter case, however, is not necessarily the case when 'η_T' is at the highest data-rate supported by PHY. It might be because of the reason that the current value of 'η_T' is not suitable for delay sensitive applications (when the MAC layer delay exceeds a threshold); therefore, lowering the transmission-rate by one step in the supported data-rate values of an 802.11 standard should be done, e.g. if the 'η_T' is at 11 Mbps, lowering one step will cause 'η_T' to be at 5.5 Mbps. In both of these cases, 'η_T' is known and the values of γ, δ, ξ are also known; therefore, equation-3.16 can easily be evaluated to provide the recommended value of 'η_{HL}'.

3.12.7 Moving Average

The output of the rate-matching block is fed into the moving average block which calculates the moving average for 'm' frames. The default value of 'm' is 30; increasing this value will decrease the responsiveness of CLRA to changes in higher layers traffic requirements. The outcome of this block serves as a limit for the statistics based rate adaptation module.

3.12.8 Statistics based Rate Adaptation

The rate adaptation block selects a certain transmission rate based on the outcome of previous transmissions. It increases the transmission rate after every 5 successful transmissions. After increasing it to a higher value if the transmission of the first 2 packet fails it reduces the transmission rate again. In fact it essentially works in an ARF like fashion however with a major change, where in this case the limit is set by the moving average block. In other words CLRA essentially selects transmission rates which can successfully meet higher layers traffic requirements. A complete block diagram of CLRA is shown in Figure 3-9.

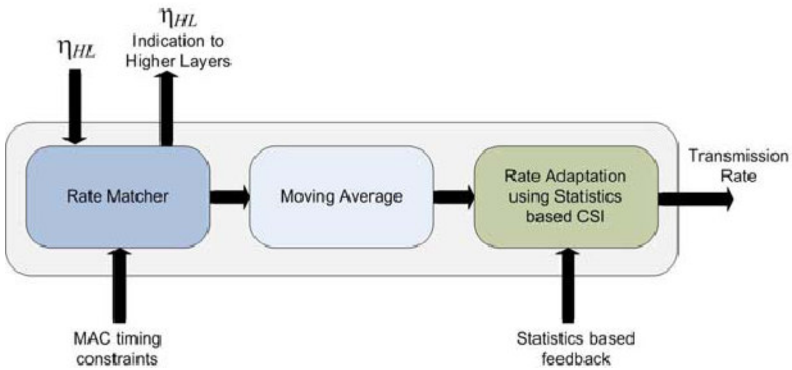

Figure 3-9: Block diagram of CLRA.

3.12.9 CLRA-loss differentiation

A frame loss-differentiation mechanism is an integral part of a rate-adaptation scheme. The loss differentiation mechanism in CLRA is based on a simple idea: which is to differentiate frame losses because of variation in channel quality from losses which occur as a result of simultaneous transmission of hidden nodes. The loss differentiation works as follows:

I. A STA maintains a frame loss window (f_{lwin}) in which it records the transmission status of the previous 'f_{lwin}' number of frames.

II. If loss ratio within an 'f_{lwin}' is greater than frame loss threshold (f_{lthr}) then enable RTS/CTS procedure for the transmission of next 'f_{lwin}' number of frames.

III. At the end of current 'f_{lwin}' (where RTS/CTS was enabled) compare the loss ratio with the loss ratio in the previous 'f_{lwin}' (with no RTS/CTS).

IV. If the loss ratio in the 'f_{lwin}' (with RTS/CTS) is more or equal to the loss ratio in the 'f_{lwin}' (without RTS/CTS) then it implies that the frame losses are because of mobility (causing loss of connectivity between the transmitter and receiver) or because of some other channel variations. Therefore, it is favourable to lower the transmission-rate of the transmitter to a one step lower value.

V. On the other hand if the loss ratio in the 'f_{lwin}' (with RTS/CTS) is less than the loss ratio in the 'f_{lwin}' (without RTS/CTS) it means that there are hidden nodes which are simultaneously transmitting and causing loss of frames. In this case, the transmission-rate remains unchanged.

Here, 'f_{lwin}' is the size of frame loss window in terms of number of frames; the default value used in simulations is four frames. Frame loss threshold (f_{lthr}) can be specified by applications however, the default value is two frames.

Frame losses because of the presence of hidden nodes can be minimized with the usage of RTS/CTS procedure but it also induces overhead in terms of MAC time. Therefore, it is important to use RTS/CTS to the minimal in order to minimize such an overhead. CLRA uses a collision-counter for deciding whether to use RTS/CTS in the next 'f_{lwin}' or send frames without RTS. The default value of the collision-counter is zero, however, when the loss ratio within an 'f_{lwin}' is greater than frame loss threshold (f_{lthr}) the collision-counter is incremented to 1. Therefore, for the next group of frames in the 'f_{lwin}' an RTS/CTS procedure will be used. However, no RTS/CTS will be used in the following 'f_{lwin}'. If the frame loss ratio in the current 'f_{lwin}' (with no RTS/CTS) is more than that in the previous 'f_{lwin}' then it implies that there are collision losses and hidden nodes are still operating. It is important to note that the collision-counter at this stage has a value of 1 which is incremented to 2. Therefore, the next two 'f_{lwin}' will have RTS/CTS frames followed by an 'f_{lwin}' without RTS/CTS. The maximum value of collision-counter used in simulations is 10 which mean that an 'f_{lwin}' without RTS/CTS is used after 10 RTS/CTS 'f_{lwin}'. The collision-counter is reset to 1 when frame loss ratio in an 'f_{lwin}' with no RTS/CTS is less than that in the 'f_{lwin}' with RTS/CTS irrespective of the current value of collision-counter. This procedure is explained in Figure 3-10.

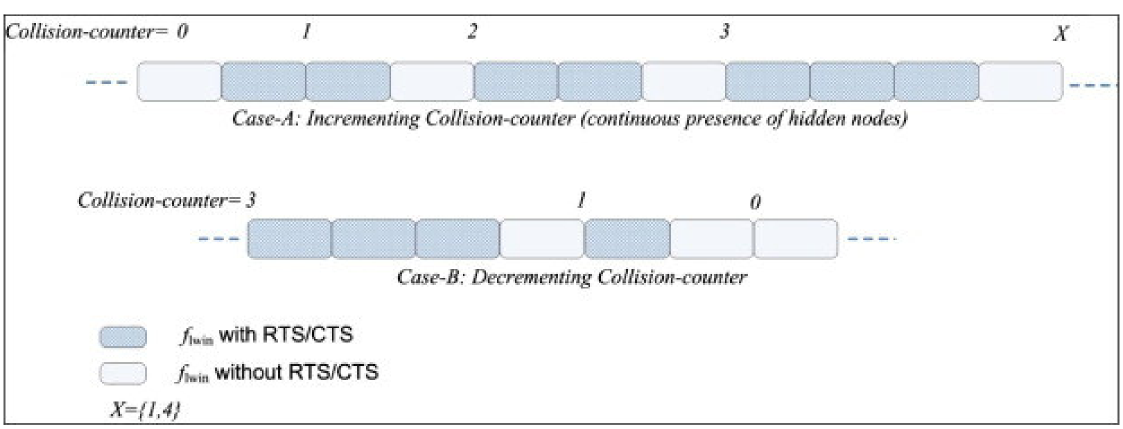

Figure 3-10: Loss differentiation using selective RTS/CTS procedure.

3.13 Rate Adaptation using Mutual Feedback (MutFed) between the IEEE 802.11 conformant transmitter-receiver pair

3.13.1 Introduction & Motivation

From the discussion in section-3.2.2, it is obvious that SNR-based rate-adaptation schemes have the ability to:

1. Quickly converge to the most suitable transmission rate, unlike the frame-loss statistics based schemes which sequentially try various transmission rates until an appropriate transmission rate is found.
2. Reflect the actual channel state and so the rate-adaptation can be performed more closely to the reality.

However, the persistent problem in all (to the best of our knowledge so far) closed-loop rate-adaptation schemes is the method through which a receiver conveys its feedback to the transmitter. The existing closed-loop rate-adaptation scheme which rely on feedback from a receiver STA require necessary changes to the original IEEE 802.11 frames specification, e.g. the RBAR [34] requires changes to the MAC data-frame, the RTS/CTS frame and PLCP header. Similarly, RAF changes the duration field in the ACK frames. Incorporating changes in the standard frames renders such rate-adaptation solutions (e.g. like the RBAR and others relying on run-time feedback from the receiver) incompatible to co-exist with the legacy (standard IEEE 802.11 conformant) STAs in a BSS.

To avoid changes to the standard MAC frames but at the same time use the advantages of SNR based schemes, the sender-side SNR based rate-adaptation mechanisms were introduced. However, various assumption are involved in designing sender-side SNR-based open loop systems, e.g. it is assumed that the SNR levels recorded for the ACK frames (at the sender side) are the equal to the SNR levels that are experienced by data-frames at a receiver.

In the following section we discuss the design of an SNR-based rate-adaptation scheme, which is closed-loop and thus rely on feedback from the receiver; however, the feedback delivery mechanism is performed without any changes to the standard MAC frames. We also introduce a frame-loss differentiation mechanism for the SNR-based rate-adaptation scheme which is different from the frame loss differentiation scheme we used for CLRA in the previous section. In the proposed rate-adaptation solution, the feedback delivery mechanism is unique, the rate-adaptation scheme itself is highly responsive and the frame-loss differentiation mechanism is extremely efficient. Each of this features of the proposed solution are discussed in the following sections.

3.13.2 Received Signal Strength measurement and representation in IEEE 802.11 standard

An IEEE 802.11 conformant transmitter uses transmit power at roughly 20 dBm (100 mW) and a standard receiver has reception sensitivity all the way to -96 dBm (2.511 x 10^{-10} mW). To represent and measure the received power at a receiver, the IEEE 802.11 standard defines an optional parameter called Received Signal Strength Indicator (RSSI). This value is measured by the PHY to represent the power observed at the antenna of a receiver while receiving the current PPDU. RSSI measurement is performed between the beginning of the *Start Frame Delimiter (SFD)*, in the PLCP preamble and the end of *Header Error Check (HEC),* in the PCLP header. The standard represents the RSSI with a 1 byte numeric value giving the RSSI an allowable range of 0 to 255. However, no vendor has reported to use the complete 0-255 range of the RSSI; so each vendor has a specific *maximum RSSI* value represented by *RSSI Max* in the standard. Cisco uses 101 different levels for the RSSI, the RSSI Max of Cisco therefore 100 [73, 74]. RSSI Max values used by Symbol and Atheros are 31 and 60 respectively.

The standard does not provide mapping between the RSSI levels and any particular power levels as measured in mW or dBm. It is left to manufacturers to map the energy values in mW to RSSI levels, decide the granularity and thus the total range of the RSSI values [73, 74]. Most of the vendors use tables for mapping the RSSI values with corresponding dBm values; the highest RSSI is usually mapped to -10 dBm or below. Any value of received power higher than -10dBm is mapped to the RSSI Max. The reason for mapping the RSSI Max to -10 dBm or lower value of the received signal power is that it is below -10 dBm that fluctuation in the received power can affect the transmission rate and other MAC functions of an 802.11. Therefore, every vendor tries to map the finite number of RSSI levels to the performance-sensitive part of the dBm curve (which starts usually at or below -10 dBm). Another justification for using such mapping is that RSSI is used for performing Clear Channel Assessment (CCA) and determination of Roaming Threshold. Both of these procedures require sensitivity of a receiver to very low energy levels which requires that there should be appropriate RSSI mapping to represent such low energy levels.

As an example of mapping the RSSI values to various dBm values by one of the vendors of IEEE 802.11 compliant wireless communication stations, Cisco [75], consider Table 3-2 shows the correlation between the dBm rating and the corresponding RSSI value for the Cisco 7920 Wireless IP Phone.

Table 3-2: Comparison of dBm and RSSI Values for Cisco 7920 Wireless IP Phones.

RSSI	5	10	15	20	25	30	35	40	45	50	55	60	65	70
dBm	-98	-97	-89	-83	-79	-75	-67	-61	-57	-49	-44	-41	-38	-34

Table 3-3, [75] shows the dBm ratings and the corresponding RSSI values for the Cisco Aironet 350 Series Access Points. The RSSI is labelled with a % sign in the Cisco Aironet Client Utility (ACU).

Table 3-3: Comparison of dBm and RSSI Values for Cisco ACU Client Adapters.

RSSI	0	5	10	15	20	25	30	35	45	50	55	60	65	70	75
dBm	-113	-108	-103	-97	-92	-87	-82	-77	-62	-58	-50	-47	-43	-39	-33

3.13.3 Receiver sensitivity

The IEEE 802.11 standard defines a receiver's minimum input level sensitivity for various transmission rates (and thus various modulation schemes). The standard specifies an upper limit of the frame error rate (FER); a receiver should experience less than 10% FER when receiving a PSDU of length 1000 octets at each of the specified levels of signal strength. Table 3-4 lists the transmission rates, modulation schemes and receiver's sensitivity as specified in the standard.

Table 3-4: IEEE 802.11 standard specification for receiver-sensitivity at various transmission rates.

Data rate in Mbps	Modulation	Minimum receiver sensitivity in dBm
2	DQPSK	-80
11	CCK	-76
24	16-QAM	-74
36	16-QAM	-70
48	64-QAM	-66
54	64-QAM	-65

3.13.4 Rate adaptation using Mutual Feedback between a transmitter and receiver pair (MutFed): *protocol operation*

The MutFed rate-adaptation scheme is essentially a closed-loop, SNR based scheme. In MutFed, a transmitter's rate-selection decision relies on the feedback of a receiver. During a rate-up process a receiver selects a transmission rate for a transmitter based on the SNR values of received frames and provides it as a feedback to the transmitter. A transmitter either approves or disapproves the feedback using its local (frame-delivery success or failure) statistics. Similarly, in the case of rate-down process, a transmitter selects a (lower) transmission rate and a receiver approves or disapproves the transmitter decision using a feedback.

3.13.4.1 Rate-up procedure

In the MutFed protocol, a rate-up action is performed in the following manner: a transmitter initiates communication at 11 Mbps with a particular receiver. At the receiving end, a receiver records the SNR levels of frames that it receives from a transmitter. The SNR values are recorded for a set of ten frames and after reception of every tenth frame, the receiver calculates the average value of SNR. The average value of SNR is used to select an appropriate transmission rate for the transmitter. In our implementation the selection of transmission rate is based on a table lookup; where the transmission rate selected by a receiver is most appropriate at the given average SNR. The receiver then informs the transmitter about the selected transmission rate by sending the 10^{th} ACK, DATA+ACK frames at the selected transmission rate. The approved transmission rate (which is essentially) the rate at which the ACK is received; is used by the transmitter for next 10 frames. A flow chart explaining the transmitter and receiver side flowcharts are given in Figure 3-11 and Figure 3-12 respectively.

It is important to highlight the mandatory rules set by the Standard for devising the rate-adaptation scheme, specifically the use of transmission rate for sending ACK frames. In a normal frame exchange, a transmitting STA needs to inform other STAs in a BSS about the duration of medium usage in a frame exchange by using the duration field in frames. This value in the duration field includes the time for transmission of a frame from the transmitting STA, the inter-frame spaces, and the time that a corresponding ACK frame would take. In this process, a transmitting STA knows about everything apart from the duration of ACK/CTS; because it is up to the receiving STA to select a certain transmission rate for sending the ACK/CTS frame. Therefore, to enable the transmitting STA to calculate the value of the duration field, the standard mandates that a receiving STA (which would send an ACK/CTS frame) should send an ACK/CTS frame at the highest rate in the BSSBasicRateSet which is less than or equal to the rate at which the transmitting STA sent the latest frame in the frame exchange sequence (which can be RTS or data frame).

However, if the transmitting STA sent the frame which is not in the BSSBasicRateSet and thus the condition outlined above could not be met, then a receiving STA should send the ACK/CTS frame at the highest mandatory rate of the PHY which is less than or equal to the rate at which the transmitting STA sent a frame.

The rules mandated by the standard for devising a rate-adaptation strategy essentially means that transmissions should only be done at rates which could be successfully decoded at the receiver, and at the same time the selection of transmission rates should not disrupt the distributed MAC operations in the BSS. In MutFed, the transmitter, for tenth frame transmission, uses a lower transmission-rate value for calculating the duration field of the frame transmission. This is done for a reason, that after the tenth frame transmission the receiver may approve a transmission-rate which is lower than the current rate. In that case, the frame exchange sequence would take longer than the duration indicated by the transmitter. For example, a transmitter sending the tenth frame (after which it expects a feedback from the receiver), at a transmission rate of 11 Mbps, would use 5.5 Mbps for receiving the ACK frame and then indicate the total duration in the frame. If the channel conditions have improved during the current ten frame window, the receiver would send ACK at a higher rate than the 11 Mbps, but on the other hand, it may suggest 5.5 Mbps by sending the ACK at 5.5 Mbps. Therefore, the MutFed transmitter station takes it into account while calculating the duration field. This essentially restricts MutFed's ability to a one step lower rate feedback. For example, if the transmission rate used in the current frame window is 11 Mbps, the receiver can suggest only 5.5 Mbps, in the case when channel conditions deteriorate; it cannot suggest 2 Mbps.

The effect of sending the ACK frames at a higher rate would result in end of communication earlier than as set in the duration field of the data frame sent by the transmitting station. The issue is elaborated further in the section-3.13.4.3.

3.13.4.2 Rate-down procedure

A transmitter switches down to a lower transmission rate after encountering two consecutive frame failures. When a transmitter retransmits a frame at a lower rate, the receiver discovers the previous frame failures that the transmitter just encountered by comparing:

$$rate_{rcvd} == rate_{current}$$

The '$rate_{rcvd}$' and '$rate_{current}$' are two variables maintained by a receiver for each (corresponding) transmitter. The '$rate_{rcvd}$' contains the transmission rate at which the transmitter sent the latest frame and the '$rate_{current}$' contains the rate which was in use of the transmitter during the current ten-frame window. For example, assume that a transmitter is operating at 5.5 Mbps during the current ten-frame window when it encountered frame failures. At the receiver the '$rate_{current}$' would contain 5.5 Mbps. After two consecutive failures the transmitter would lower the transmission rate to 2 Mbps; when the receiver receives the retransmitted frame at a lower rate, the '$rate_{rcvd}$' would have 2 Mbps.

In normal conditions, a receiver provides feedback after tenth-frame, but in the case when:

$$rate_{rcvd} \mathrel{!=} rate_{current}$$

(Which implies that the transmitter lowered the transmission rate); the immediately following frame (ACK, Data+ACK) is used for feedback.

The receiver's feedback just after transmission failures suggests either of two options:

a) If the transmission rate suggested by the SNR of latest frame is less than $rate_{current}$ then it means that retransmissions were caused due to actual deterioration of channel quality. In this case, the feedback frame is transmitted at $rate_{rcvd}$ (which in our previous example is the 2 Mbps transmission rate). Feedback from the receiver at a lower rate implies that the transmitter's decision to lower the rate is correct and the receiver affirms that decision. In this situation, the transmitter keeps the newly selected (lower, 2 Mbps) rate for the next ten-frame window.

b) While if the SNR values of latest frame suggests a rate equal to $rate_{current}$ then it implies that the retransmissions are likely to be caused by collisions due to the presence of hidden stations. The feedback frame is transmitted at $rate_{current}$ (which is equal to 5.5 Mbps in our example); as an indication to the transmitter that the decision to lower the rate is incorrect. After receiving the feedback frame at a higher rate, a transmitter immediately switches back to the (higher, 5.5 in the example) rate specified in the feedback.

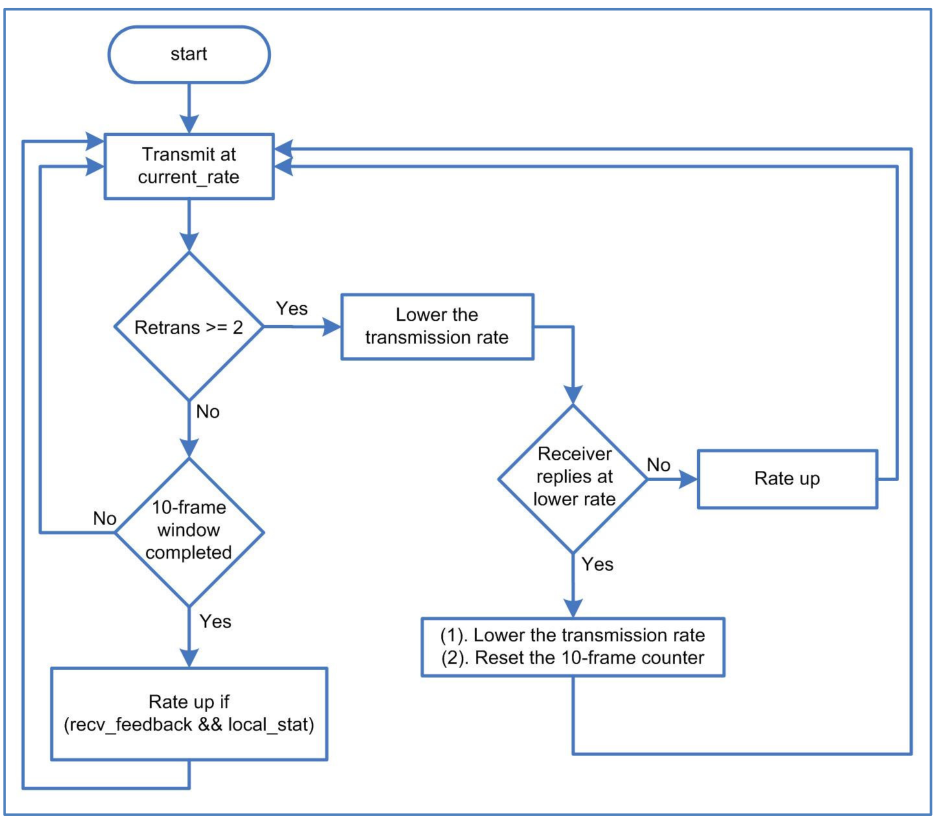

Figure 3-11: MutFed's operation flowchart at the transmitter side.

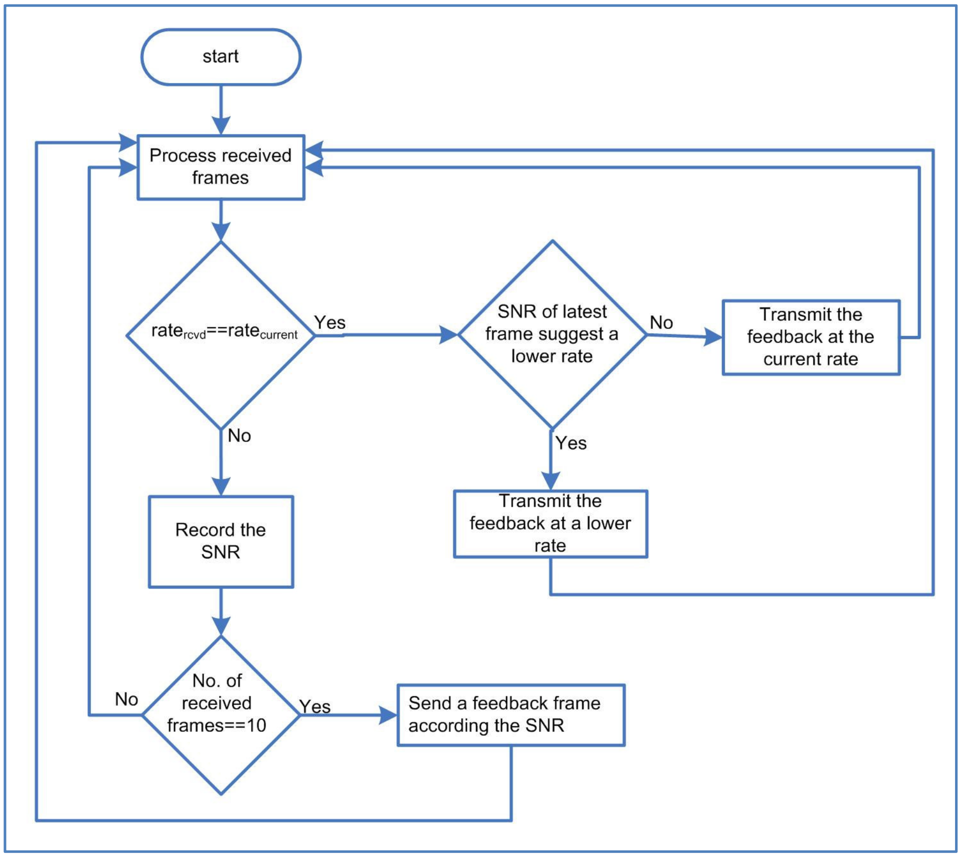

Figure 3-12: MutFed's operation flowchart at the receiver side.

3.13.4.3 Frequency of feedback

The essential purpose of providing feedback to the transmitter is to improve responsiveness of the rate-adaptation scheme when the channel quality improves. The frequency of feedback thus plays an important role for determining the overall responsiveness of a rate-adaptation scheme.

In case of MutFed the default feedback interval is after transmission of 10 frames. Various other feedback interval are used for examining the effects of variable feedback interval on the performance of MutFed in chapter-4.

For a highly responsive rate-adaptation scheme ideally every ACK or Data+ACK frame should convey the feedback to the transmitter. However, per-frame feedback is not recommended for two reasons:

1. Highly frequent feedback can be misleading at times, e.g. at times of high SNR fluctuations, an improved channel quality indicated by a receiver would not last for

longer for a transmitter to react and change its transmission rate accordingly. Therefore, per-frame SNR feedback may not reflect the true measure of the channel quality.

2. Secondly and more importantly, during a frame transmission procedure, an 802.11 conformant station indicates the duration of medium used with the help of the 'duration' field. This duration includes the time taken by the data frame at the current transmission rate, the SIFSs and the duration which an ACK frame would take from the receiver to the source for acknowledging the frame transmission. In case of the MutFed, when a feedback is provided by the receiver, i.e. the ACK is sent a higher rate, the frame exchange sequence ends quickly than the time indicated to the other stations. Just after the frame exchange is over, all stations would contend for the medium. For the source station, its Network Allocation Vector (NAV) is not set while for other stations; their NAV is set to the duration of frame exchange time. Therefore, the source station would find the medium to be idle earlier than other stations.

This process introduces unfairness in the medium access procedure. In case of per-frame feedback, this condition the degree of unfairness would be very high. A solution to overcome the unfairness in medium access for the source station, is to program the source station to set a timer (similar to NAV) and assign it a value equal to the duration field. The source station should not activate its CCA before the expiry of that timer. Therefore, all stations would activate their CCA mechanisms at the same time, as indicated in the duration field.

3.13.4.4 MutFed- loss differentiation mechanism

It is important to note that in the MutFed protocol, a receiver coordinates with a transmitter to identify the actual cause of retransmissions. The feedback provided by a receiver immediately after transmission failures performs MAC level loss differentiation.

As shown in Figure 3-13, after reception of frame-B, the receiver compares its SNR with the SNR average that it has for the current 10-frame window. If the there is not a significant difference (where the received SNR equals the receiver sensitivity threshold of a lower rate), in the SNR of the latest frame (frame-B) and the average SNR of the current 10-frame window; it simply suggests that the previous frame losses that the transmitter experienced were because of collisions. Therefore, the transmitter should not lower the transmission rate.

As shown in Figure 3-13, the receiver immediately informs the transmitter about its feedback by sending ACK frame at $Rate_x$ (frame 'D' in the Figure 3-13). On the other hand if there is a significant difference in the SNR, the receiver sends the ACK frame at a lower rate 'Rate(x-1)' (frame 'C' in the Figure 3-13). Therefore, frame-losses can be immediately identified and a quick feedback be delivered to the transmitter.

The loss differentiation mechanism used by MutFed turns out to be highly responsive when compared with selective RTS/CTS [52 - 59], RTS/CTS window MAC level loss differentiation mechanisms.

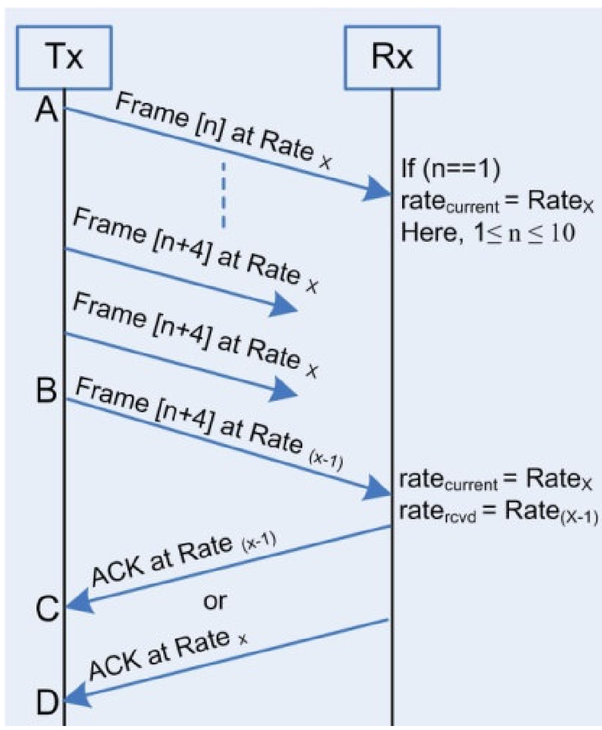

Figure 3-13: Frame exchange between a transmitter and receiver. Frame-A, is start of 10-frame window, the receiver always reset the $rate_{current}$ to the rate at which it receives first frame in a window. Frame-B, a third retransmission at $Rate_{(x-1)}$ after 2 frame losses at $Rate_x$, Frame-C is an ACK/Data+ACK from receiver at a lower rate than the $rate_{current}$. Frame-D is an ACK/Data+ACK from receiver at the same rate as $rate_{current}$..

3.14 Summary

This chapter presented two rate-adaptation schemes. One of the proposed rate adaptation scheme works at the sender/transmitter side, and determines the CSI with the help of successful or failed transmissions of frames. However, the proposed scheme uses a cross-layer communication frame work, to incorporate PHY and application specific detail while performing rate-adaptation. An on-demand incremental strategy is used to increase the highest allowable transmission rate the demand of the applications or as a requirement because of the prevailing constraints of the underlying layers. A frame loss differentiation mechanism using an RTS/CTS has also been proposed to distinguish the reasons for various types of frame losses. The loss differentiation mechanism enables the rate-adaptation to work more intelligently.

The second contribution of this chapter is another rate-adaptation scheme called MutFed, which is essentially an SNR based approach. In MutFed, a receiver works in close coordination with a transmitter and delivers a feedback with the help of ACK frames, with the need for modification in the standard frames. Using the same feedback mechanism a very efficient frame-loss differentiation mechanism is also introduced in MutFed. The proposed protocol is highly responsive and rate-adaptation scheme quickly reacts to variations in channel quality.

Both of the proposed schemes, reduces the number of retransmissions, which mostly occur because of selection of an inappropriate transmission rate. By reducing the total number of retransmissions, a high degree of communication efficiency in terms of lower transmission delay, lower medium occupancy, spectral efficiency and considerably higher cumulative throughput can be achieved; this is further explored in the following chapter.

3.15 References

[1]. "IEEE Standard for Information technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," ,IEEE Std 802.11-2007 (Revision of IEEE Std 802.11-1999) C1-1184, 2007.

[2]. D. Yang, T. Lee, K. Jang, J. Chang, S. Choi, "Performance enhancement of multirate IEEE 802.11 WLANs with geographically scattered stations", IEEE Transactions on Mobile Computing, vol.5, no.7, pp. 906-919, July 2006

[3]. T. Joshi, A. Mukherjee, Y. Yoo, D. P. Agrawal, "Airtime Fairness for IEEE 802.11 Multirate Networks,", IEEE Transactions on Mobile Computing, vol.7, no.4, pp.513-527, April 2008

[4]. S. Khan, S. A. Mahmud, H. S. Al-Raweshidy, "Rate Adaptation for Multi-Rate Wireless Stations: A Cognitive Approach", ICT-Mobile Summit, 2008.

[5]. Kamerman and L. Monteban, "WaveLAN®-II: a high-performance wireless LAN for the unlicensed band," Bell Labs Technical Journal, vol. 10, pp. 119, 1997.

[6]. M. Lacage, M. H. Manshaei and T. Turletti, "IEEE 802.11 rate adaptation: a practical approach," Proceedings of the 7th ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems, pp. 126-134, 2004.

[7]. J.C. Bicket, Bit-rate selection in wireless networks, M.S. thesis, MIT, Feb 2005.

[8]. Multiband Atheros Driver for WiFi (MADWiFi), Available: http://madwifi-project.org/

[9]. B. Sadeghi, V. Kanodia, A. Sabharwal, and E. W. Knightly, OAR: An Opportunistic Auto-Rate Media Access Protocol for Ad Hoc Networks, Wireless Networks, vol. 11, no. 1-2, pp. 39-53, 2005.

[10]. M., Ashraf, A., Jayasuriya, "Improved Opportunistic Auto Rate protocols for wireless networks," Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on , vol., no., pp.1-6, 15-18 Sept. 2008.

[11]. W. Jianfeng, Z. Hongqiang, F. Yuguang, M.C. Yuang, "Opportunistic media access control and rate adaptation for wireless ad hoc networks," *Communications, 2004 IEEE International Conference on* , vol.1, no., pp. 154-158 Vol.1, 20-24 June 2004.

[12]. T., Kim, J.T., Lim, "Rate Adaptation Based on Collision Probability for IEEE 802.11 WLANs" IEICE TRANSACTIONS on Communications, 2008.

[13]. C., Jaehyuk, P., Kihong, K., Chong-kwon, "Cross-Layer Analysis of Rate Adaptation, DCF and TCP in Multi-Rate WLANs," *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE* , vol., no., pp.1055-1063, 6-12 May 2007.

[14]. T. Joshi, D., Ahuja, D., Singh, D. P. Agrawal, "SARA: Stochastic Automata Rate Adaptation for IEEE 802.11 Networks," *Parallel and Distributed Systems, IEEE Transactions on* , vol.19, no.11, pp.1579-1590, Nov. 2008.

[15]. A., Singh, D., Starobinski, "A Semi Markov-based Analysis of Rate Adaptation Algorithms in Wireless LANs," *Sensor, Mesh and Ad Hoc Communications and Networks, 2007. SECON '07. 4th Annual IEEE Communications Society Conference on* , vol., no., pp.371-380, 18-21 June 2007.

[16]. K. Ramachandran, H. Kremo, M. Gruteser, P. Spasojevic, and I. Seskar. Scalability analysis of rate adaptation techniques in congested ieee 802.11 networks: An orbit testbed comparative study. In WoWMoM, 2007.

[17]. K. Jong-Ok, H. Tode, K. Murakami, "Service-based rate adaptation architecture for IEEE 802.11e QoS networks," *Global Telecommunications Conference, 2005. GLOBECOM '05. IEEE* , vol.6, no., pp.5 pp.-3345, 2-2 Dec. 2005.

[18]. Q., Daji, C., Sunghyun, K.G., Shin, "Goodput analysis and link adaptation for IEEE 802.11a wireless LANs," *Mobile Computing, IEEE Transactions on* , vol.1, no.4, pp. 278-292, Oct-Dec 2002.

[19]. L. Shaohe, W. Xiaodong, Z. Xingming, L. Chi, "Efficient Rate Adaptation in Multi-rate 802.11 Networks," *Computers and Communications, 2007. ISCC 2007. 12th IEEE Symposium on* , vol., no., pp.545-550, 1-4 July 2007.

[20]. J. P. Pavon, C. Sunghyun, "Link adaptation strategy for IEEE 802.11 WLAN via received signal strength measurement," *Communications, 2003. ICC '03. IEEE International Conference on* , vol.2, no., pp.1108-1113 vol.2, 2003

[21]. K.K., Leung, P.F., Driessen, K., Chawla, Q., Xiaoxin, "Link adaptation and power control for streaming services in EGPRS wireless networks," *Selected Areas in Communications, IEEE Journal on* , vol.19, no.10, pp.2029-2039, 2001.

[22]. M. H., Manshaei, T., Turletti, M., Krunz, "A media-oriented transmission mode selection in 802.11 wireless LANs," *Wireless Communications and Networking Conference, 2004. WCNC. 2004 IEEE* , vol.2, no., pp. 1228-1233 Vol.2, 21-25. 2004.

[23]. Won. Jong-ho, K. Chong-kwon. A Downlink Rate Adaptation Scheme in IEEE 802.11 WLANs using Overhearing, International Conference on Information Networking, ICOIN. pp 1-5. 2008.

[24]. J., H., Yun, "Reliable rate adaptation for block acknowledgment in IEEE 802.11e wireless LANs," *Electronics Letters* , vol.44, no.3, pp.216-217, January 31 2008

[25]. C. Jaehyuk, N, Jongkeun , P. Kihong, K. Chong-kwon Kim, "Adaptive Optimization of Rate Adaptation Algorithms in Multi-Rate WLANs," Network Protocols, 2007. ICNP 2007. IEEE International Conference on , vol., no., pp.144-153, 16-19 Oct. 2007

[26]. T. Ozcelebi, M. Oguz Sunay, T.A. Murat, M.C. Reha, "Cross-Layer Optimized Rate Adaptation and Scheduling for Multiple-User Wireless Video Streaming," *Selected Areas in Communications, IEEE Journal on* , vol.25, no.4, pp.760-769, May 2007

[27]. I. Haratcherev, J. Taal, K. Langendoen, R. Lagendijk, and H. Sips. Automatic IEEE 802.11 rate control for streaming applications. Wirel. Commun. Mob. Comput. pp. 421-437. 2005.

[28]. J.L.C. Wu, L. Hung-Huan, L.Yi-Jen, "An adaptive multirate IEEE 802.11 wireless LAN," *Information Networking, 2001. Proceedings. 15th International Conference on* , vol., no., pp.411-418, 2001

[29]. I. Tinnirello, C. Sunghyun, "Temporal fairness provisioning in multi-rate contention-based 802.11e WLANs," *World of Wireless Mobile and Multimedia Networks, 2005. WoWMoM 2005. Sixth IEEE International Symposium on a* , vol., no., pp. 220-230, 13-16 June 2005.

[30]. S. Biaz, W. Shaoen, "Rate adaptation algorithms for IEEE 802.11 networks: A survey and comparison," *Computers and Communications, 2008. ISCC 2008. IEEE Symposium on* , vol., no., pp.130-136, 6-9 July 2008.

[31]. K., Seongkwan, L., Sung-Ju, C., Sunghyun, "The impact of IEEE 802.11 MAC strategies on multi-hop wireless mesh networks," *Wireless Mesh Networks, 2006. WiMesh 2006. 2nd IEEE Workshop on* , vol., no., pp.38-47, 25-28 Sept. 2006.

[32]. I. Haratcherev, K. Langendoen, R. Lagendijk, and H. Sips. Hybrid rate control for IEEE 802.11. In *Proceedings of the Second international Workshop on Mobility Management &Amp; Wireless Access Protocols*. 2004.

[33]. K.-Y. Doo, J. young Song, and D.-H. Cho. Enhanced transmission mode selection in IEEE 802.11a LAN system. In *Proceedings of Vehicular Technology Conference*, pages 5059–5062, Sept. 2004.

[34]. G. Holland, N. Vaidya, and P. Bahl. A rate-adaptive MAC protocol for multi-hop wireless networks. *Mobile Computing and Networking*, pp. 236-251, 2001.

[35]. C. Chun-cheng, L. Haiyun, S., Eunsoo, N.H., Vaidya, W., Xudong, "Rate-Adaptive Framing for Interfered Wireless Networks," *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE* , vol., no., pp.1325-1333, 6-12 May 2007.

[36]. Z. Jiansong, K., Tan, Z., Jun, W. Haitao, Z. Yongguang, "A Practical SNR-Guided Rate Adaptation," *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE* , vol., no., pp.2083-2091, 13-18 April 2008.

[37]. Z. Li, A. Das, A. K. Gupta, Nandi, S., "Full auto rate MAC protocol for wireless ad hoc networks," *Communications, IEE Proceedings-* , vol.152, no.3, pp. 311-319, 3 June 2005.

[38]. W., Li-chun, Y., Kuang-nan, C. Ming-bing, Y., Yu-ren, H, Ping-jung, “Cross-layer rate adaptation mechanism for WLAN”, United States Patent 7423970. 09, September 2008.

[39]. X., Qiuyan, M., Hamdi, "Smart sender: a practical rate adaptation algorithm for multirate IEEE 802.11 WLANs," *Wireless Communications, IEEE Transactions on* , vol.7, no.5, pp.1764-1775, May 2008

[40]. S. Biaz, W. Shaoen, "Loss Differentiated Rate Adaptation in Wireless Networks," *Wireless Communications and Networking Conference, 2008. WCNC 2008. IEEE* , vol., no., pp.1639-1644, March 31 2008-April 3 2008.

[41]. I. Haratcherev, K. Langendoen, I. Lagendijk, and H. Sips. Application-Directed Automatic 802.11 Rate Control: Design Rationale and Preliminary Results. In ASCI 2003, pages 56–60, Heijen, The Netherlands, June 2003.

[42]. X., Qiuyan, M., Hamdi, H. Tsz Ho, “Practical Rate Adaptation for IEEE 802.11 WLANs”. GLOBECOM, pp.1-5. 2006.

[43]. P. Chevillat, J. Jelitto, A.N. Barreto, H. L.Truong, "A dynamic link adaptation algorithm for IEEE 802.11 a wireless LANs," *Communications, 2003. ICC '03. IEEE International Conference on* , pp. 1141-1145 vol.2, 11-15 May 2003.

[44]. Q. Daji, C., Sunghyun, "Fast-responsive link adaptation for IEEE 802.11 WLANs," *Communications, 2005. ICC 2005. 2005 IEEE International Conference on* , vol.5, no., pp. 3583-3588 Vol. 5, 16-20 May 2005.

[45]. C. Hoffmann, H. Manshaei, T. Turletti, "CLARA: closed-loop adaptive rate allocation for IEEE 802.11 wireless LANs," *Wireless Networks, Communications and Mobile Computing, 2005 International Conference on* , vol.1, no., pp. 668-673 vol.1, 13-16 June 2005

[46]. P.A.K. Acharya, A., Sharma, E.M., Belding, K.C., Almeroth, K., Papagiannaki, "Congestion-Aware Rate Adaptation in Wireless Networks: A Measurement-Driven Approach," *Sensor, Mesh and Ad Hoc Communications and Networks, 2008. SECON '08. 5th Annual IEEE Communications Society Conference on* , vol., no., pp.1-9, 16-20 June 2008

[47]. M. Rodrig, C. Reis, R. Mahajan, D. Wetherall, and J. Zahorjan. Measurement-based Characterization of 802.11 in a Hotspot Setting. In ACM SIGCOMM E-WIND Workshop, 2005.

[48]. A. Jardosh, K. Ramachandran, K. Almeroth, and E. Belding-Royer. Understanding Link-Layer Behavior in Highly Congested IEEE 802.11b Wireless Networks. In Proceedings of ACM SIGCOMM Workshop E-WIND, Philadelphia, PA, August 2005.

[49]. A. Jardosh, K. Ramachandran, K. Almeroth, and E. Belding. Understanding Congestion in IEEE 802.11b Wireless Networks. In ACM/USENIX International Measurement Conference, Berkeley, CA, October 2005.

[50]. W., Shao-Cheng, A., Helmy, "BEWARE: Background traffic-aware rate adaptation for IEEE 802.11," *World of Wireless, Mobile and Multimedia Networks, 2008.*

WoWMoM 2008. 2008 International Symposium on a , vol., no., pp.1-12, 23-26 June 2008.

[51]. N., Baldo, F., Maguolo, S., Merlin, A., Zanella, M., Zorzi, D., Melpignano, D., Siorpaes, "GORA: Goodput Optimal Rate Adaptation for 802.11 Using Medium Status Estimation," *Communications, 2008. ICC '08. IEEE International Conference on* , vol., no., pp.4916-4921, 19-23 May 2008.

[52]. J., Kim, S., Kim, S., Choi, D., Qiao, "CARA: Collision-Aware Rate Adaptation for IEEE 802.11 WLANs," *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings* , vol., no., pp.1-11, April 2006.

[53]. K. Seongkwan, C. Sunghyun, Q. Daji and K. Jongseok , “Enhanced Rate Adaptation Schemes with Collision Awareness”, lecture notes, Springer Berlin / Heidelberg, vol. 4479/2007, pp. 1179-1182 . 2007.

[54]. S. H.,Wong, H., Yang, S., Lu, and V. Bharghavan. Robust rate adaptation for 802.11 wireless networks. In *Proceedings of the 12th Annual international Conference on Mobile Computing and Networking* . MobiCom '06. pp 146-157. 2006.

[55]. F., Maguolo, M., Lacage, T., Turletti, "Efficient collision detection for auto rate fallback algorithm," *Computers and Communications, 2008. ISCC 2008. IEEE Symposium on* , vol., no., pp.25-30, 6-9 July 2008.

[56]. C. Xi, Q. Daji, Y. Jeonggyun, C. Sunghyun, "Probabilistic-Based Rate Adaptation for IEEE 802.11 WLANs," Global Telecommunications Conference, 2007. GLOBECOM '07. IEEE , vol., no., pp.4904-4908, 26-30 Nov. 2007.

[57]. Q. Pang, S. C. Liew, V. C. M. Leung, “Design of an Effective Loss- Distinguishable MAC Protocol for 802.11 WLAN”, IEEE Communications Letters, vol.9, no.9, pp. 781- 783, Sep 2005.

[58]. Q. Pang, V.C.M. Leung, S. C., Liew. "A rate adaptation algorithm for IEEE 802.11 WLANs based on MAC-layer loss differentiation," *Broadband Networks, 2005. BroadNets 2005. 2nd International Conference on* , pp. 659-667 Vol. 1, 3-7 Oct. 2005.

[59]. S. Kim, S. Choi, D. Qiao, and J. Kim, “Enhanced Rate Adaptation Schemes with Collision Awareness”.In Proc. IFIP Networking'07, Atlanta, GA, May 14~18, 2007.

[60]. J.L.C. Wu, L. Hung-Huan, L.Yi-Jen, "An adaptive multirate IEEE 802.11 wireless LAN," *Information Networking, 2001. Proceedings. 15th International Conference on* , vol., no., pp.411-418, 2001

[61]. I. Tinnirello, C. Sunghyun, "Temporal fairness provisioning in multi-rate contention-based 802.11e WLANs," *World of Wireless Mobile and Multimedia Networks, 2005. WoWMoM 2005. Sixth IEEE International Symposium on a* , vol., no., pp. 220-230, 13-16 June 2005.

[62]. D., Aguayo, J., Bicket, S., Biswas, G., Judd, and R., Morris. "Link-level measurements from an 802.11b mesh network". *SIGCOMM Comput. Commun. Rev.* 34, 4. pp. 121-132. 2004.

[63]. S. Garg, M. Kappes, "An experimental study of throughput for UDP and VoIP traffic in IEEE 802.11b networks," Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE , vol.3, no., pp.1748-1753 vol.3, 20-20 March 2003.

[64]. J., Hakyung, K., Taekyoung, C., Yanghee, S., Yongho, "A Scalable Rate Adaptation Mechanism for IEEE 802.11e Wireless," *Future generation communication and networking (fgcn 2007)* , vol.1, no., pp.505-509, 6-8 Dec. 2007.

[65]. W. Kun, Y. Fan, Z., Qian, D.O., Wu, Y., Xu, "Distributed Cooperative Rate Adaptation for Energy Efficiency in IEEE 802.11-Based Multihop Networks," *Vehicular Technology, IEEE Transactions on* , vol.56, no.2, pp.888-898, March 2007.

[66]. M. Zawodniok, J. Sarangapani, "Energy-efficient rate adaptation MAC protocol for ad hoc wireless networks," *24th IEEE International Conference* on *Performance, Computing, and Communications, 2005. IPCCC 2005.*, vol., no., pp. 389-394, 7-9 April 2005.

[67]. S. Khan, S. A. Mahmud, H.S., Al-Raweshidy, "A Cross-Layer Solution for Dynamic Selection of Data-Rate in WLAN using Traffic Differentiation," 16th IST Mobile and Wireless Communications Summit, vol., no., pp.1-5, 1-5 July 2007.

[68]. S. Shakkottai, T. Rappaport and P. Karlsson, "Cross-layer design for wireless networks," Communications Magazine, IEEE, vol. 41, pp. 74-80, 2003.

[69]. V. Srivastava and M. Motani, "Cross-layer design: a survey and the road ahead," Communications Magazine, IEEE, vol. 43, pp. 112-119, 2005.

[70]. M. van Der Schaar, N. Sai Shankar, "Cross-layer wireless multimedia transmission: challenges, principles, and new paradigms," Wireless Communications, IEEE , vol.12, no.4, pp. 50-58, Aug. 2005

[71]. N. Nang, R. Sankar and Jungsik Lee, "Improving ad hoc network performance using cross-layer information processing," pp. 2764-2768, 2005.

[72]. V. Kawadia and P. Kumar, "A cautionary perspective on cross-layer design," Wireless Communications, IEEE [See also IEEE Personal Communications], vol. 12, pp. 3-11, 2005.

[73]. J. Bardwell. Converting signal strength percentage to dbm values. Technical report, WildPackets, November 2002.

[74]. J. Bardwell. ""You Believe You Understand What You Think I Said…" The Truth About 802.11 Signal And Noise Metrics". Available (19-January-2009): http://madwifi-project.org/attachment/wiki/UserDocs/RSSI/you_believe_D100201.pdf?format=raw

[75]. Cisco reference. Available online (09-March-2009): http://www.cisco.com/en/US/docs/voice_ip_comm/cuipph/7920/5_0/english/design/guide/wrlradio.html

Chapter 4

Design of a Rate-adaptation protocol for multi-rate capable PHY- *Performance Tests and Comparative Analysis*

4.1 Introduction

In chapter-3 we proposed two rate-adaptation schemes called CLRA and MutFed, each belonging to a different class of rate-adaptation schemes. The responsibility of a rate-adaptation scheme is to acquire channel state information and then select the most appropriate transmission rate for the future frame transmissions. There are several features of a rate-adaptation scheme which have to be considered and each of such features has its own implications on the overall communication efficiency. For instance, a rate-adaptation scheme should quickly detect the true CSI (and thus should be responsive to variation in channel quality), it should minimize the chances of frame failures during the rate-adaptation procedures (and thus minimize the retransmissions and related spectrum, energy, and delay) and at the same time it should be capable to differentiate between the actual causes of frame losses.

This chapter discusses various performance tests and comparative analysis through a number of simulation scenarios for CLRA and MutFed by comparing their performance mainly with ARF and AARF.

4.2 Performance tests and comparative analysis of CLRA

4.2.1 CLRA simulation model

The simulation model which is used to model the CLRA is incorporated in the model used for IEEE 802.11 MAC protocol, called *wlan_mac,* in OPNET [1]. The wlan_mac is implemented with the help of a state-transition diagram.

In a state-transition diagram a transition from one state to another is made based on fulfilment of a specific condition. The condition which is required for a transition from one state to another state is defined in a Header block e.g. when a frame is successfully received by a station, this event should trigger the MAC process model so that appropriate functions/code (which reside in another state) can be executed. Such triggers are usually set with the help of condition definitions.

With every transition from one state to another one, usually there are certain functions which are needed to be called during the transition, such functions are called *transition executives*.

There are two type of states, in the state-transition modelling: a forced state (marked green) and an unforced sate (marked red). A forced state is essentially a code holder, and after the code within the forced state is executed, another transition is made out of this state. This is used generally for clear/easy definition of the process models. On the other hand, the process model remains in the unforced state unless a certain condition is fulfilled. There are two parts in a state: the '*enter executive'* part, where the resident code is executed whenever a transition is made into that particular state and another part called '*exit executive',* which contains the code which is executed just before leaving the current state.

In order to test the performance of the cross-layer rate-adaptation (CLRA) scheme, a simulation model was implemented in OPNET, as shown in Figure 4-1. The CLRA model is based on the following equation, which was derived in chapter-3:

$$\eta_T = \frac{\eta_{HL}\delta_x}{\delta_x - \eta_{HL}(\gamma_x + \varepsilon_x)} \tag{4.1}$$

The above equation essentially estimates the transmission rate which is required to be used by a transmitter at a given time, by using: the traffic arrival rate from higher layers, the medium access delays and the frame size.

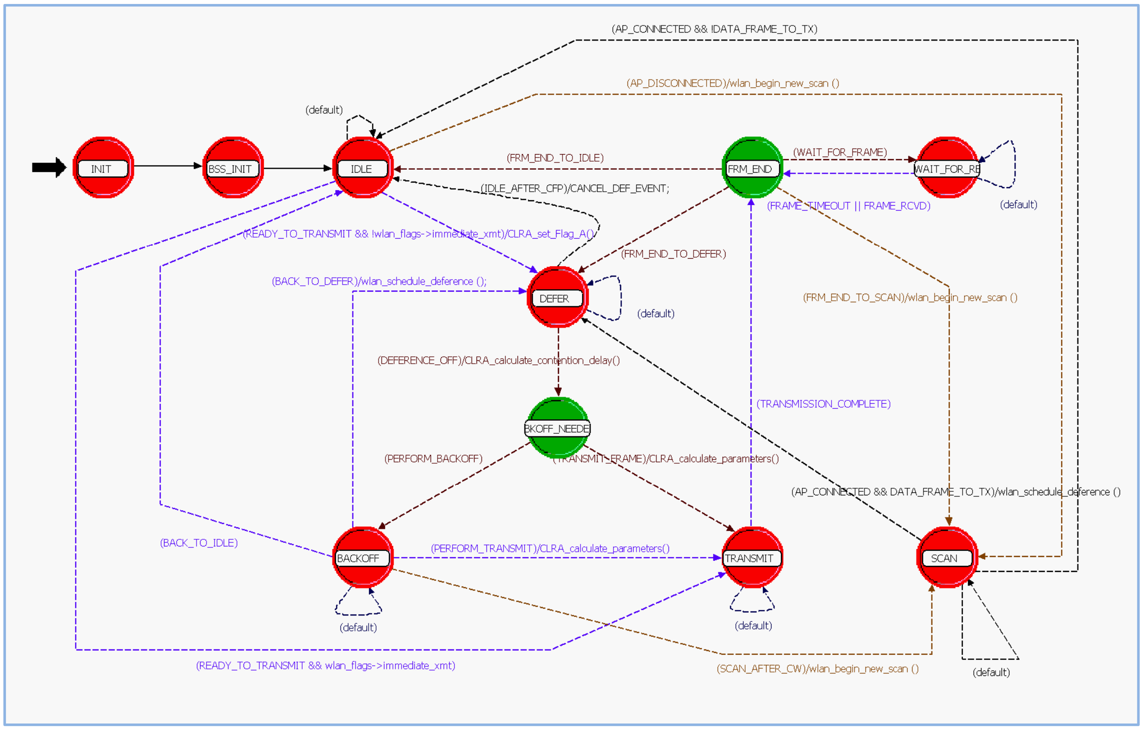

Figure 4-1: State transition diagram of the IEEE 802.11 MAC model and CLRA in OPNET.

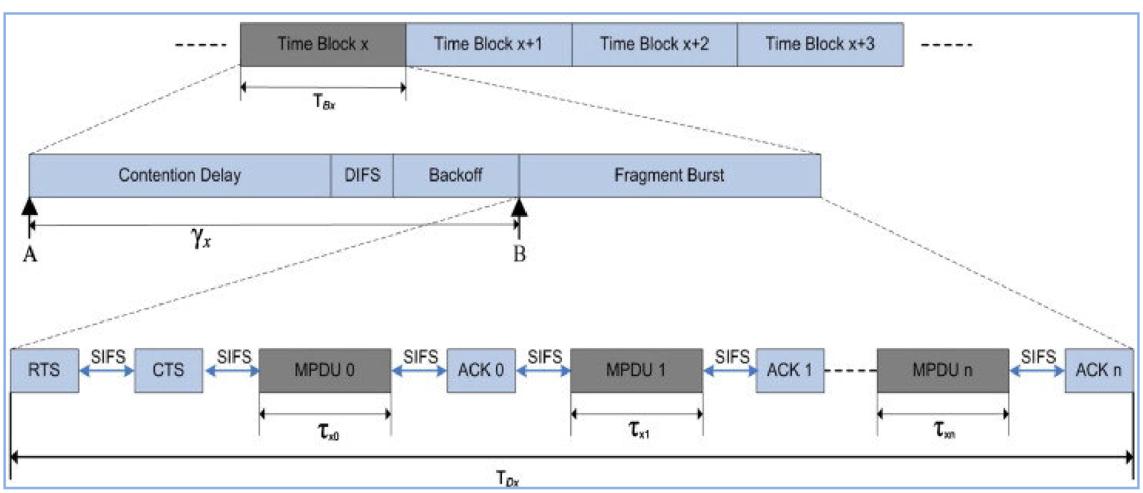

Figure 4-2: Timing diagram for IEEE 802.11 MAC operation.

The function in the Figure 4-1, called *CLRA_calculate_contention_delay()*, is invoked for calculating the contention delay experienced by a frame while waiting for transmission. The function calculates contention delay by using a *flag* which marks the start and end of the parameter 'γ' as indicated in the Figure 4-2. The flag is set whenever the backoff counter is frozen as a result of medium usage by another station. This is done at the time when the *wlan_schedule_deference()* function is called; the function *wlan_schedule_deference()* itself is called when the condition *BACK_To_DEFER* is true. Therefore, the processing done by *CLRA_calculate_contention_delay()* is simplified. An excerpt from the code is given in Figure 4-3 which shows how the contention delay is calculated by the CLRA model. Here, the variables: *CLRA_n_DIFS, CLRA_flag_A* and *CLRA_contention_delay* are *state variables*. The function *op_sim_time()* returns the current simulation time.

```
Static void CLRA_calculate_contention_delay (void)
{.
.
++CLRA_n_DIFS;
CLRA_contention_delay +=op_sim_time () – CLRA_flag_A;
}
```

Figure 4-3: CLRA_calculate_contention_delay function.

The backoff delay is calculated as: *CLRA_backoff_delay= backoff_slots * slot_time;*

Here, *CLRA_backoff_delay*, *backoff_slots* and *slot_time* are all *state variables*. The *backoff_slots* shows the value of the backoff counter for the current frame transmission, multiplied by the duration of a slot-time gives the total time spent as a result of the backoff process.

At a time when the medium is idle for DIFS interval and the backoff counter reaches zero, the MAC is clear to transmit; therefore, in the CLRA model, the function *CLRA_calculate_parameters()* is called. In this function, the variable, 'γ_x', used in the equation-4.1 is calculated using:

*CLRA_gamma= CLRA_contention_delay + CLRA_n_DIFS * difs_time + CLRA_backoff_delay;*

The parameters $'\delta'$ is the frame size, which is obtained by calling an OPNET specific function for calculating the size of a packet (*op_pk_total_size_get()*) and adding the return value to a constant called *WLANC_MPDU_HEADER_SIZE.* For calculating $'\varepsilon'$, a function called *CLRA_calculate_sigma()* is used. The main code for calculating sigma is given in Figure 4-4. Here, the variables: *control_data_rate,* and *sifs_time* are state variables.

```
Static double CLRA calculate sigma (int n, Boolean RTS CTS)
{
....
If(RTS CTS)
        Sigma    (1   +   2*n)   *   sifs time   +   n   *   TXTIME CTRL   (WLAN ACK LENGTH)   +
        TXTIME CTRL DR(WLAN CTS LENGTH,   control data rate)   +   TXTIME CTRL DR   (WLANC RTS LENGTH,
        control data rate);
else
        sigma   (2*n -1) * sifs time + n* TXTIME CTRL(WLANC ACK LENGTH);
FRET(sigma);
}
```

Figure 4-4: Function for calculatingδ.

For, calculating the limiting value of the transmission rate (i.e. the η_T), the parameters $\delta, \gamma_x, \varepsilon_x$ and η_{HL} are required. The value of η_{HL} is calculated from the average value of 10 previous time-blocks.

The η_T thus obtained using the equation-4.1 is then fed into a moving average block (which is also a function). The output of the moving average block defines the limiting value for the statistics-based rate-adaptation scheme used in the CLRA.

To analyze various performance parameters relating to rate adaptation, a number of scenarios with varying device density, traffic rates and mobility patterns were simulated. Likewise, special cases of channel variations because of hidden node problems are also investigated and its effect on rate adaptation schemes is discussed.

4.2.2 Backoff delay variations

In order to simulate higher layer's traffic, we generated variable-bit rate (VBR) traffic with a maximum bit rate of 1 Mbps. Likewise; we modelled scenarios with varying device density (2, 10, 25 and 35 nodes) for the purpose of analyzing the effect of device density on performance of the proposed algorithm.

In Figure 4-5 the backoff delay variations is shown in various scenarios with a different degree of device density. The backoff delay is considerably lower for scenarios having 2 and 10 nodes only. However, it increases as the number of nodes is increased further. For reasons discussed in chapter-3, device density in has almost a direct relationship with the probability of two or more stations attempting to use the medium simultaneously. Higher backoff delay implies that the duration of γ_x within a time block (T_B) is increased and the rate adaptation algorithm needs to shrink T_{Dx} within the PHY limits in order to meet the higher layers' traffic requirements.

In other words, if γ_x is longer enough within a particular time block, it means that added amount of data will arrive from higher layers before the transmitter gets a turn to transmit on the medium. If the amount of data, arriving at a particular rate is higher than the MSDU size within a time block, the algorithm will operate at the highest possible transmission rate determined by a PHY standard. However, equation-4.1 will not be valid and higher layer data needs to be buffered.

If this situation prevails for some time, it may result in packet losses in case of buffer overflows.

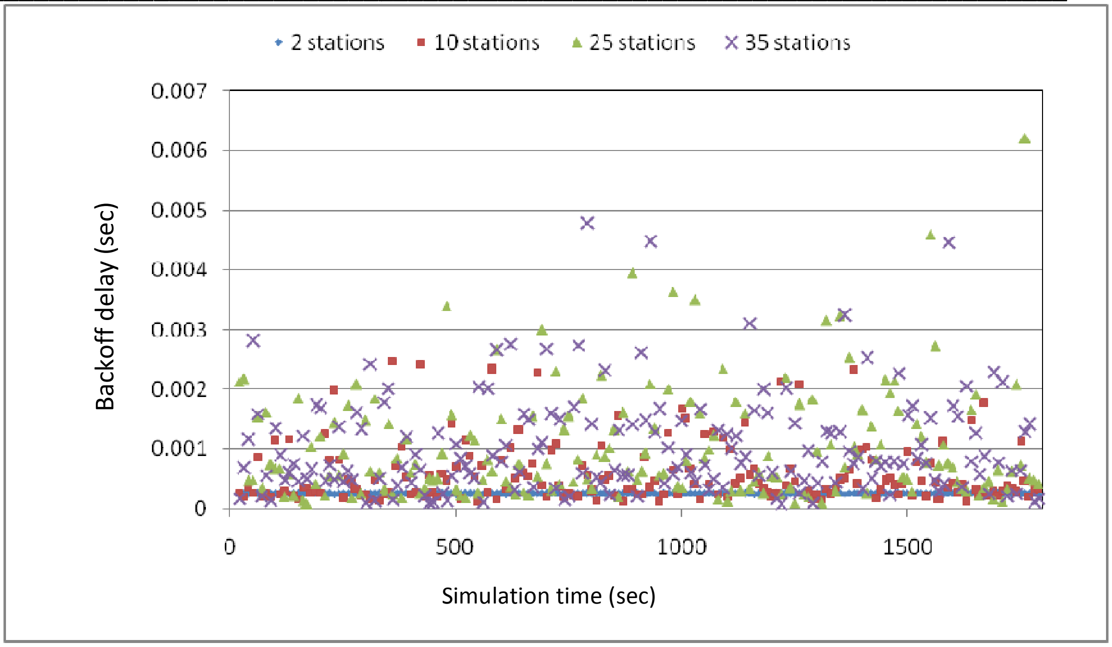

Figure 4-5: Backoff delay increases with increase in the number of stations in a BSS. Scenarios with 2 stations are compared with three other scenarios with 10, 25 and 35 stations. The reason for higher backoff delay is that the probability of simultaneous medium access by more than one stations increases with increase in the number of stations.

4.2.3 Contention delay variations

In the Figure 4-6, the contention delay experienced by stations in scenarios with 2, 10, 25 and 35 wireless stations is shown. It is obvious that as the number of nodes increases, medium usage is increased as well and therefore, other nodes have to contend for longer time until the medium becomes free and find their turn to transmit. Like, backoff delay, it also contributes to values of γ_x. Variations in γ_x due to either or both of these factors will imply that the transmission rate should vary accordingly, even if there is no variation in the higher layers' traffic rates.

A comparison of η_{HL} (Higher layer's traffic), η_T (Effective data-rate of transmitter according to equation-4.1) and the operational data-rate of the transmitter (because the transmitter can only select from a discrete set of values) in a scenario with only source and destination nodes is shown in Figure 4-7. It is obvious from Figure 4-5 and Figure 4-6 that in case of two nodes, the contention and backoff delays are considerably lower. Even then in order to match the higher layers' traffic rate, the transmitter should be operating at slightly higher rate. The red marks at the top of blue lines show the difference in data rate which is required to cover up for the time wasted in contention and backoff process.

As the wasted time within every time block increases because of longer contention and backoff delays, this difference in data rate mounts higher and higher till it reaches the maximum data-rate supported by a corresponding PHY.

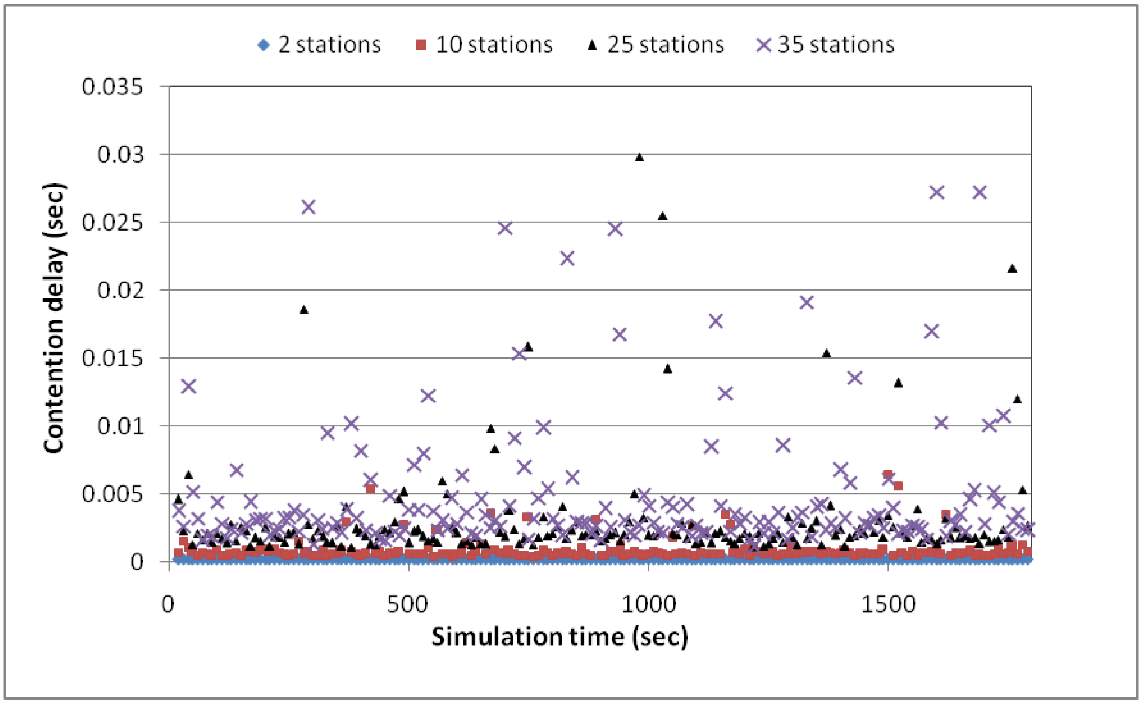

Figure 4-6: Medium contention delay increases with increase in the number of stations in a BSS; it also depends on the traffic patterns and the transmission rates used by the stations. Scenarios with two stations are compared with three other scenarios with 10, 25 and 35 stations. The stations used CBR traffic and used a fixed transmission rate.

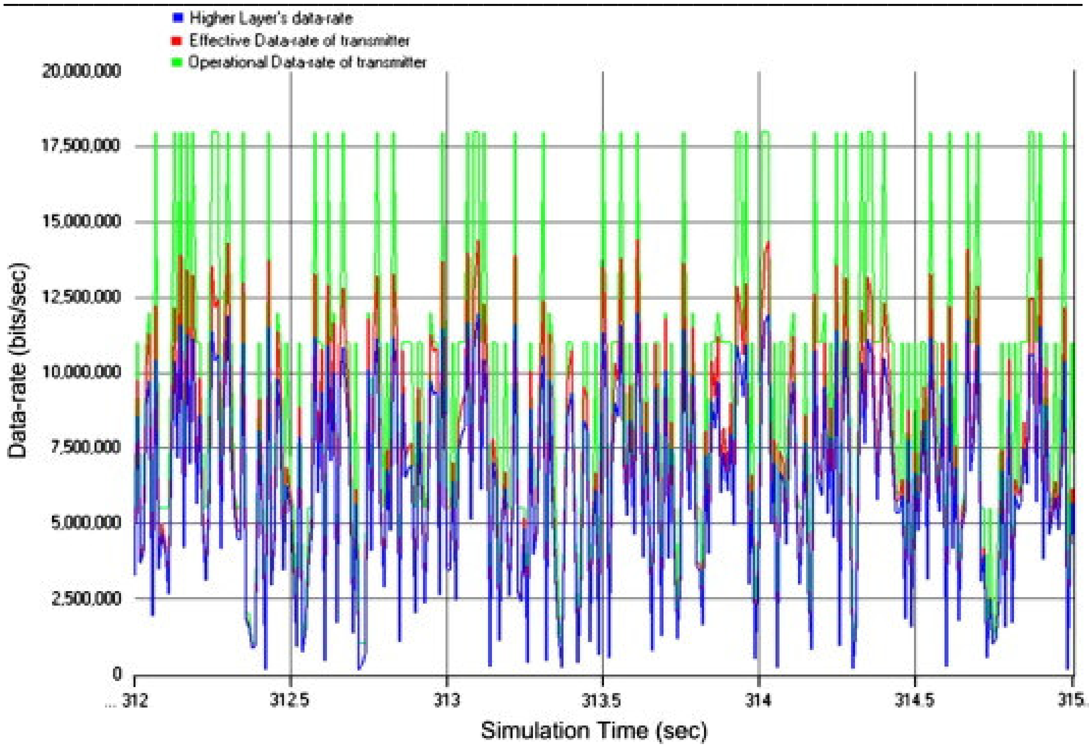

Figure 4-7: Comparison of η_{HL}, η_T and operational data-rate of a transmitter in a scenario with two stations.

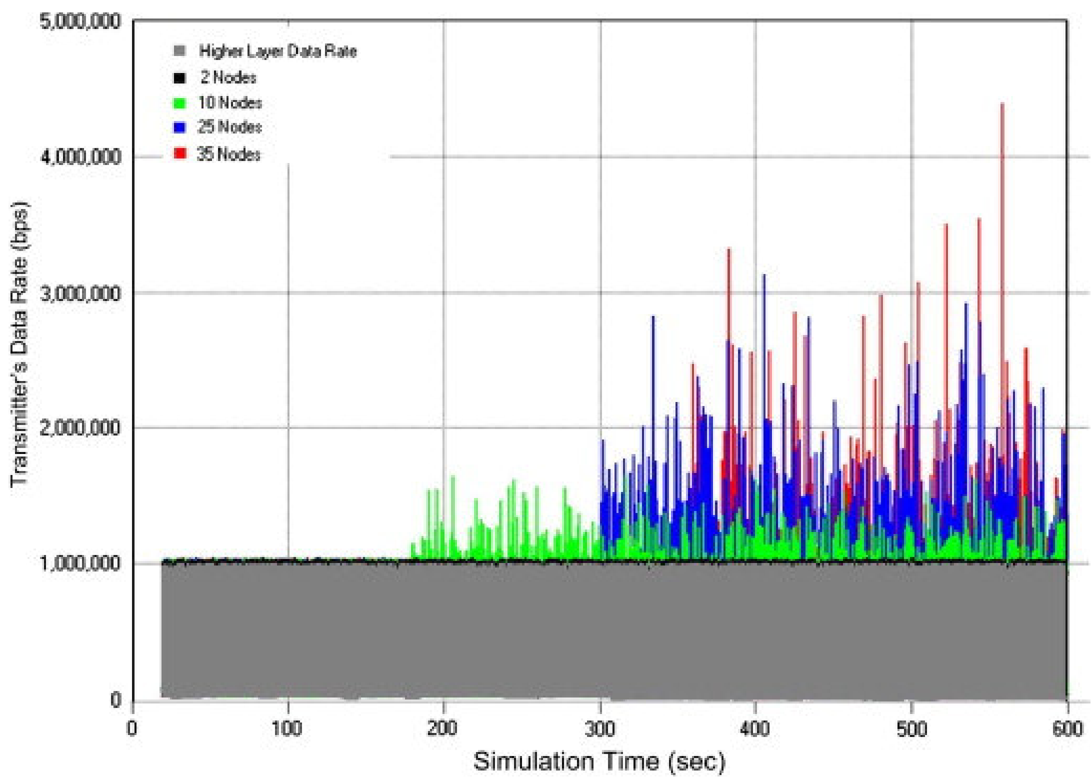

Figure 4-8: Transmission-rate selection in four scenarios with station density varying from 2 to 10, 25 and 35 stations.

A similar situation where the values of η_T from various scenarios having different number of nodes are plotted together with the η_{HL} values in Figure 4-8. It is evident that for the same higher layers' traffic requirements, the transmission-rate should go higher if the number of nodes increases. This difference is lower in case of two nodes but considerably increases as the number of nodes increases. For number of nodes higher than 15 up to 35 (as shown in the simulation results), the transmission-rate should be between the range of 2–4 times of the data-rate requirements of higher layers.

From these results it can be assumed that even though the transmitter may be able to operate at higher data-rates (up to 54 Mbps), however, because of the MAC layer timing constraints (contention and backoff delays) the effective data-rate supported by the MAC layer is never equivalent to the highest value. At times of higher contention delays, the effective data-rate might be reduced to half, one third or one fourth of the highest data-rate supported by a PHY. When this situation arises, such that the application requirements cannot be fulfilled by the underlying MAC protocol there is a need for signalling this information to the application layer. Such information can be conveyed through a cross layer interface as mentioned in the previous chapter. The application layer can further utilize it to take necessary action.

It is important to note that Figure 4-8 only shows the actual values of η_{HL} and η_T as given by equation-4.1. However, in practice the values of η_T are rounded up to the nearest values in the data rate array supported by a particular PHY e.g. values of 1.3 Mbps as given by equation-4.1 are rounded up to 2, 3.4 Mbps are rounded up to 5.5 Mbps, etc.

4.2.4 Average value of η_T

It is important to highlight that the outcome of the equation-4.1 is dependent on a number of factors and as a result the instantaneous value of the η_T is highly variable. Therefore, to stabilize the limiting rate value, we feed the instantaneous value of η_T to a moving average block.

The moving-average block gets an input from the rate-matcher, which in turn determines the runtime requirements of the higher layers along with the consideration of the timing constraints of underlying MAC protocol. The output of moving-average block acts as a limiter for the statistical rate determination block. Therefore, contrary to ARF [2], and

AARF [3]; CLRA will only select transmission rates marked by the moving-average block where as ARF and AARF will go as high as the highest allowable transmission rate by the PHY. It is important to note that even in very simple scenarios when the application's outbound traffic requirements are minimal and where the distance between the transmitter and receiver is such that the transmissions at high data rates cannot reach the destination, ARF and AARF inherently try to select the highest transmission rate repeatedly, which results in retransmissions.

4.2.5 CLRA comparison with existing rate-adaptation schemes

Real-time applications are delay sensitive and excessive MAC layer retransmissions because of frame losses penalize their performance. However, most of the rate adaptation algorithms inherently try to operate at higher transmission-rate. In scenarios where the transmitter moves away from the destination (or vice versa) or is positioned at a distance where higher rate transmission could not be possible; such a behaviour (of always trying to select a higher transmission rate) can cause severe retransmissions.

To analyze this effect we modelled a scenario where: a destination node starts moving exactly after 1 min to a position almost 200 m away from the source station. At that position it stays for about 3 min and then moves further by 140 m to another position where it stops till the end of simulation period. A comparison of the number of retransmissions per second in three different runs of the same scenario modelling ARF, AARF and without a rate adaptation scheme is shown in Figure 4-9.

It is quite clear that when the distance between the source and destination increases then transmissions at made at a higher transmission rate have fewer chances to reach the destination stations. In case of ARF, AARF and SampleRate [4], a higher transmission rate is selected after a predefined number of frame transmissions. In many cases, such a higher rate transmission may not even be necessary. In the scenario under consideration, the ARF and AARF repeatedly tried to increase the transmission rate which caused relatively higher retransmissions. The number of retransmissions is comparatively lower in the case of AARF because of the backoff after successive frame failures. The transmission-rate values selected in this scenario by the ARF algorithm are given in Figure 4-10.

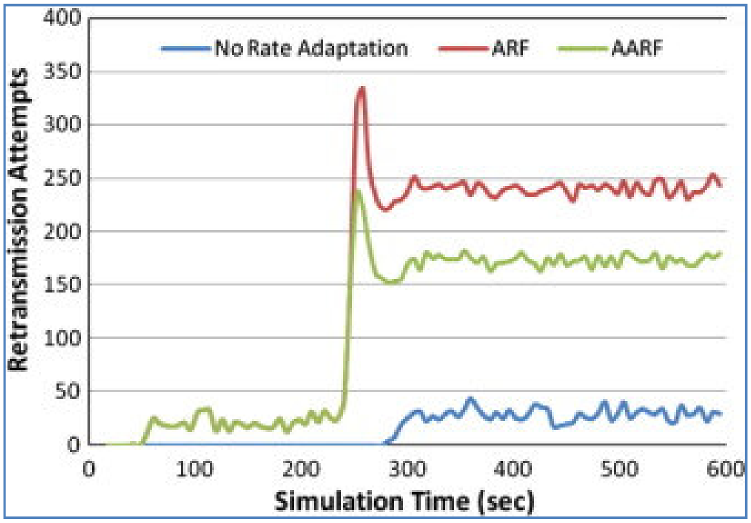

Figure 4-9: Analysis of retransmissions in various scenarios (ARF, AARF and no rate adaptation).

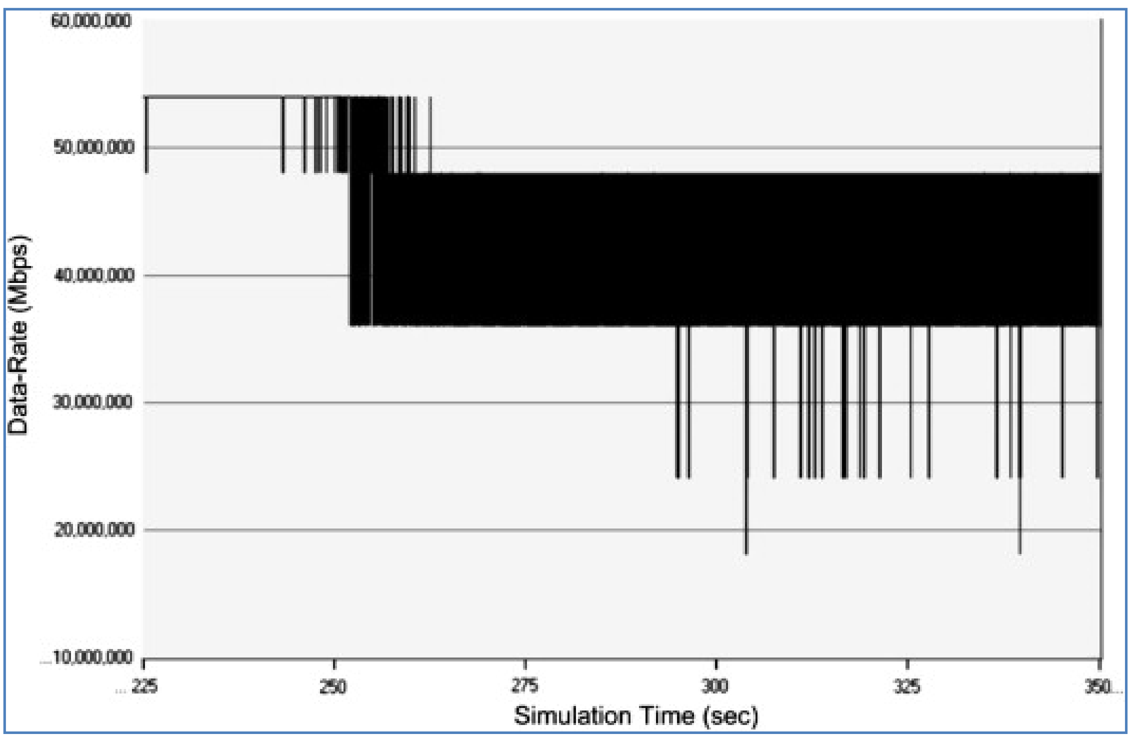

Figure 4-10: Transmission rates selection using ARF in the simulation scenario.

An important observation which can be made from this scenario is that it is not always optimal to operate at the highest transmission-rate. In the previous scenario, the application requirements could have been met even by operating at a much lower transmission-rate. That's why the number of retransmission was comparatively lower when the station operated at a constant 11 Mbps (in the no-rate-adaptation scenario).

Therefore, the CLRA algorithm which closely follows the application requirements under the prevailing MAC layer timing constraints can clearly perform better in such scenarios. One might argue that operating at lower data-rates will occupy the channel for a longer time. This is true for cases when the BSS has a number of contending stations, in that case, operating at a lower data-rate will penalize rest of the contending stations. However, as stated before, CLRA adjusts the operational data-rate according to the MAC layer timing constraints. In other words, if the device density increases, an accordingly higher operational data-rate is selected.

To demonstrate this fact, we modelled a scenario with 11, IEEE 802.11g conformant wireless stations, in a 100 × 100 m area and CLRA was enabled for rate adaptation. In this scenario, one node was configured with constant bit rate (CBR) higher layer traffic of 5 Mbps. This station was configured to start communication after 20 s of simulation time. Another station was configured to start communication after 180 s with 500 Kbps traffic. In this discussion, we will refer to this station as a Light Load station. Four other stations with a higher layer traffic of 2 Mbps each was configured to start at 240, 360, 480 and 540 s of simulation time. This was followed simultaneously by four more stations at 600 s simulation time. The Figure 4-11 shows that till 480 s time, the Light Load station was transmitting at lowest possible rate. However, soon after that when new more stations started communication, the medium usage became higher resulting in longer contention. As a result, even for low traffic, the station has to operate at considerably higher data-rate to cope up with the traffic requirements of higher layers.

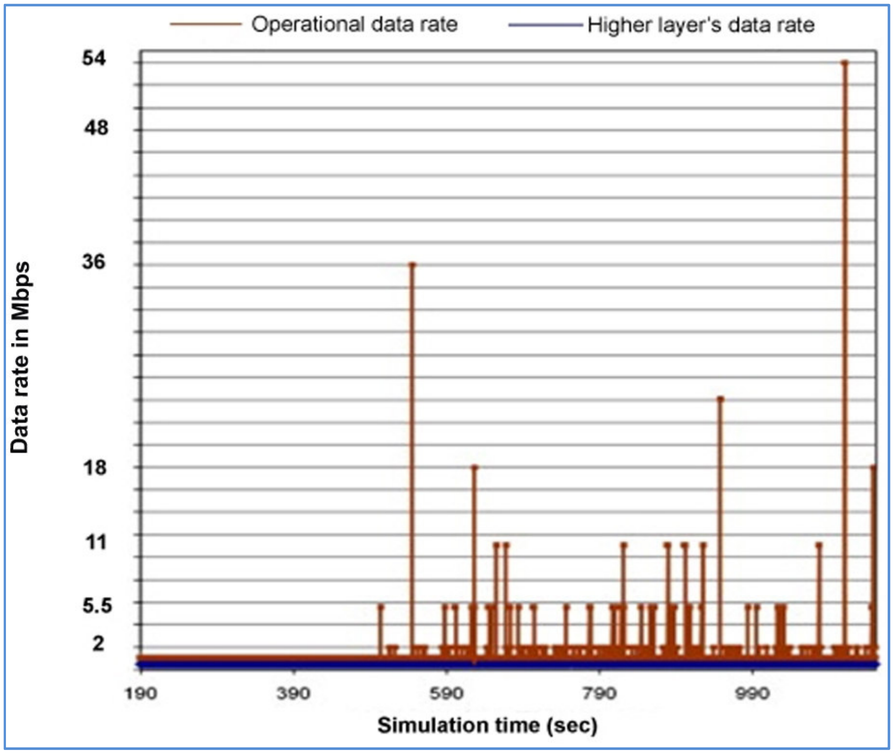

Figure 4-11: Operational rate varies when more stations in a BSS start transmission even if the higher layer's requirements remain the same.

4.2.6 CLRA loss differentiation

CLRA assess channel variations using frame loss-differentiation mechanism before performing a rate adaptation. Most of existing rate adaptation mechanisms selects a lower data-rate in case of frame losses. However, it can be misleading and can cause performance degradation if the exact reason of channel variations is not determined. To analyze a similar situation, we modelled a hidden node scenario. In this scenario, all wireless stations remained stationary during the whole simulation. Therefore, we can ignore the effects of channel variations because of node's mobility. Figure 4-12 shows throughput comparison of ARF (without loss differentiation) and CLRA using its loss differentiation. The hidden nodes are configured to start transmission (turn on during the simulation) after 3 min (180 s). Before the hidden nodes started transmission, the throughput of ARF is much higher than CLRA. However, after the frame losses started as a result of the presence of hidden nodes, the performance of ARF clearly started to suffer (after 180 s).

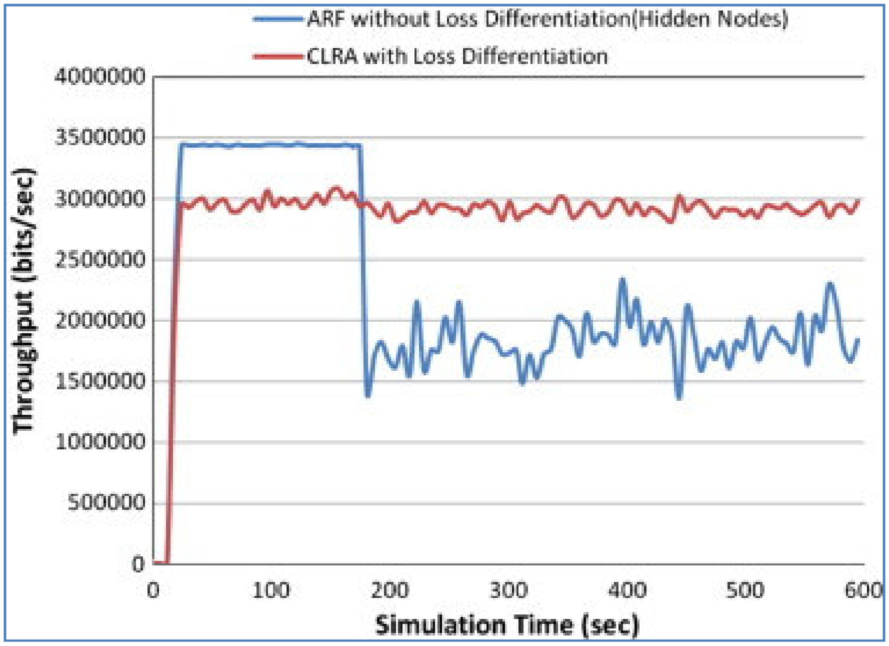

Figure 4-12: CLRA loss differentiation

In another simulation scenario 12, IEEE 802.11g ERP conformant stations were configured to use 5mW transmit power and constant frame size of 1024 bits. The stations were configured with constant bit rate (CBR) and variable bit rate (VBR) traffics with random destinations in an ad-hoc network environment. All probes for statistics collection were set up at a station, called primary station. Hidden stations were configured in the vicinity of the 'primary station' with transmission configuration setup in a way so that they caused frame losses because of simultaneous transmissions from time to time.

The results shown in Figure 4-13 reports the instantaneous selection of transmission rates by ARF, AARF and CLRA in the same scenario. In case of ARF, the transmission rate was increased after transmission of every 10 frames, which in many cases, in that scenario caused transmission failure. A similar trend is followed by AARF. Likewise, due to the lack of a loss differentiation mechanism, a simultaneous transmission from a hidden station which caused a frame loss, was also considered to be deterioration of the channel quality. And thus the transmission rate was once again lowered unnecessarily.

A time-average plot of the retransmission attempts during packet transmission in case of ARF, AARF and CLRA is shown in Figure 4-14. This statistics, in the figure are collected as a sum of 500 samples. In other words, each value represents the number of retransmissions required for transmitting 500 packets. There is no significant difference among the three schemes in terms of retransmissions especially in this scenario. AARF stays for a longer time at a lower rate after it encounters a retransmission than ARF. However, because of the persistent hidden node problem, the number of retransmissions is same for ARF and AARF which forces AARF to stay at the minimum rate than ARF and CLRA.

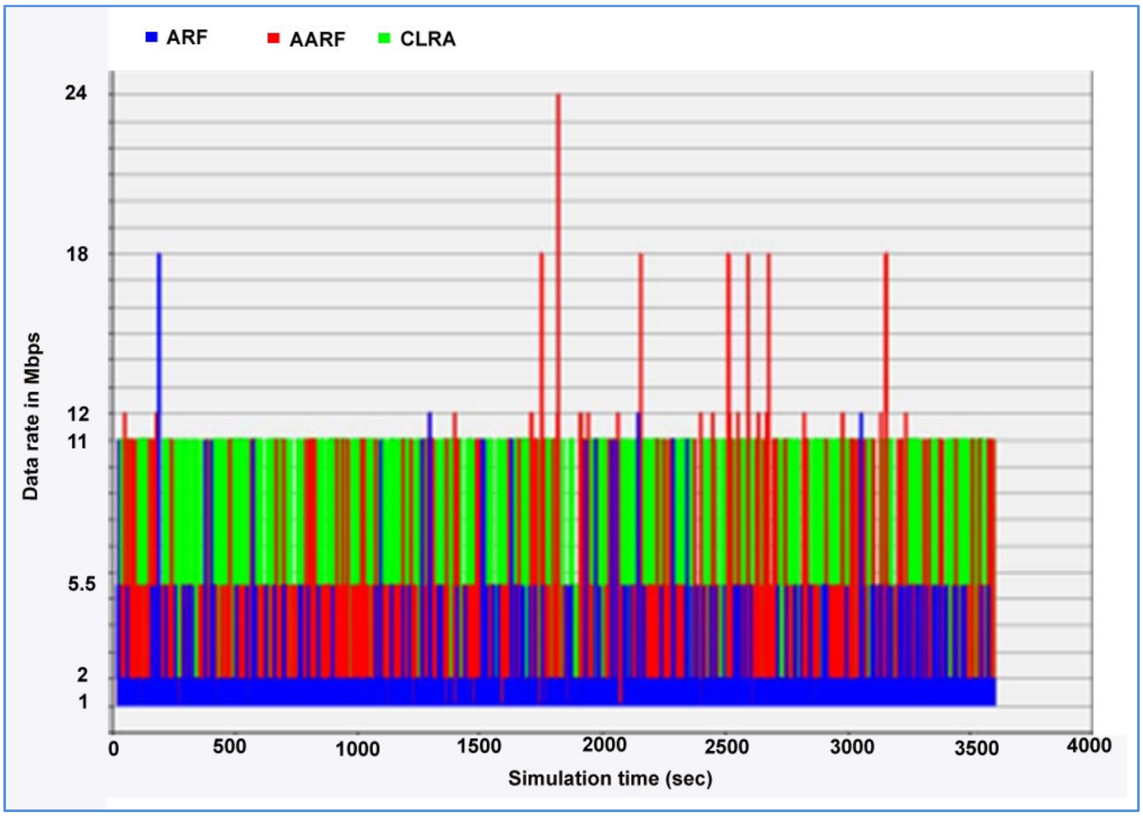

Figure 4-13: Comparison of ARF, AARF and CLRA.

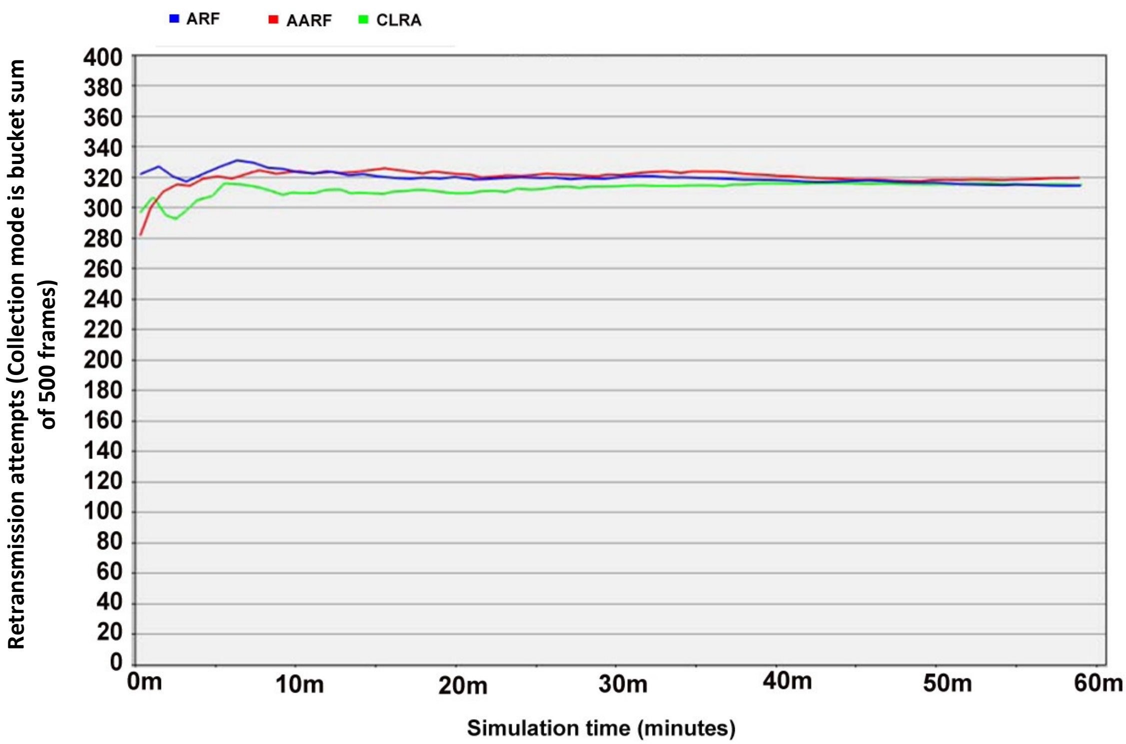

Figure 4-14: Time-Average plot of the frame retransmissions in ARF, AARF and CLRA scenarios.

4.2.7 Effects of retransmissions and role of a rate-adaptation scheme

Selection of an appropriate transmission rate is significant in determining the efficiency of a communication system. Operating at lower transmission rates causes higher medium occupancy which results in reduction of the capacity of the shared wireless network. Likewise, it also causes higher MAC level delay and higher consumption of energy.

While on the other hand, transmissions at a higher rate have a higher probability to encounter transmission failure. In IEEE 802.11 conformant stations, a frame is considered to be lost, if it is not acknowledged by the intended receiver. After transmission of frame the source station waits for duration called ACK-timeout, and if no ACK is received, the latest transmission is considered to be lost. The procedure after a frame failure is very expensive in terms of communication efficiency for instance:

- After the first transmission the source station waits for ACK-timeout duration, which is equal to: SIFS time + SlotTime + PHY-RX-START-Delay.

- The backoff counter is incremented, which means that the station would have to wait for a longer time before accessing the medium for the retransmitted frame.
- Generally, in most of the rate-adaptation schemes cited in this thesis, the retransmission is made at a lower rate than the first transmission (which failed). This means that in case of a retransmission, a station has to acquire the medium twice, for a longer duration the second time. It not only causes inefficient usage of the share wireless medium, but it also causes communication delays, lower throughput and inefficient use of limited battery powered stations' energy.

Figure 4-15 shows energy consumption comparison of these three rate adaptation schemes i.e. the ARF, AARF and CLRA in the same scenario. In order to get a meaningful value of energy consumption we calculate the energy consumed per Mega bits transmission. Here the maximum MAC protocol Data Unit (MPDU) size was kept constant at 1248 bits (where the MSDU=1024 and MPDU header=224) and transmit power=5mW. Therefore, transmission time per frame multiplied by the transmit power gives the energy consumption per frame or almost per kilo bits.

To get this statistic in terms of Joules per Mega bits the collection mode was changed to bucket/SUM 1000 values. In other words, each value plotted in the curves in Figure 4-15 is collected as a sum of values in a bucket (the bucket size is 1000). It also includes the energy consumption for retransmitted frames. It can be noticed from Figure 4-15 that CLRA clearly performs better in terms of energy efficiency. In this scenario, CLRA uses at maximum point approximately 82% and at minimum point approximately 51% of the energy consumed by ARF. On the other hand it uses approximately 51% and 41% of the energy consumed under ARF scheme at maximum and minimum points respectively.

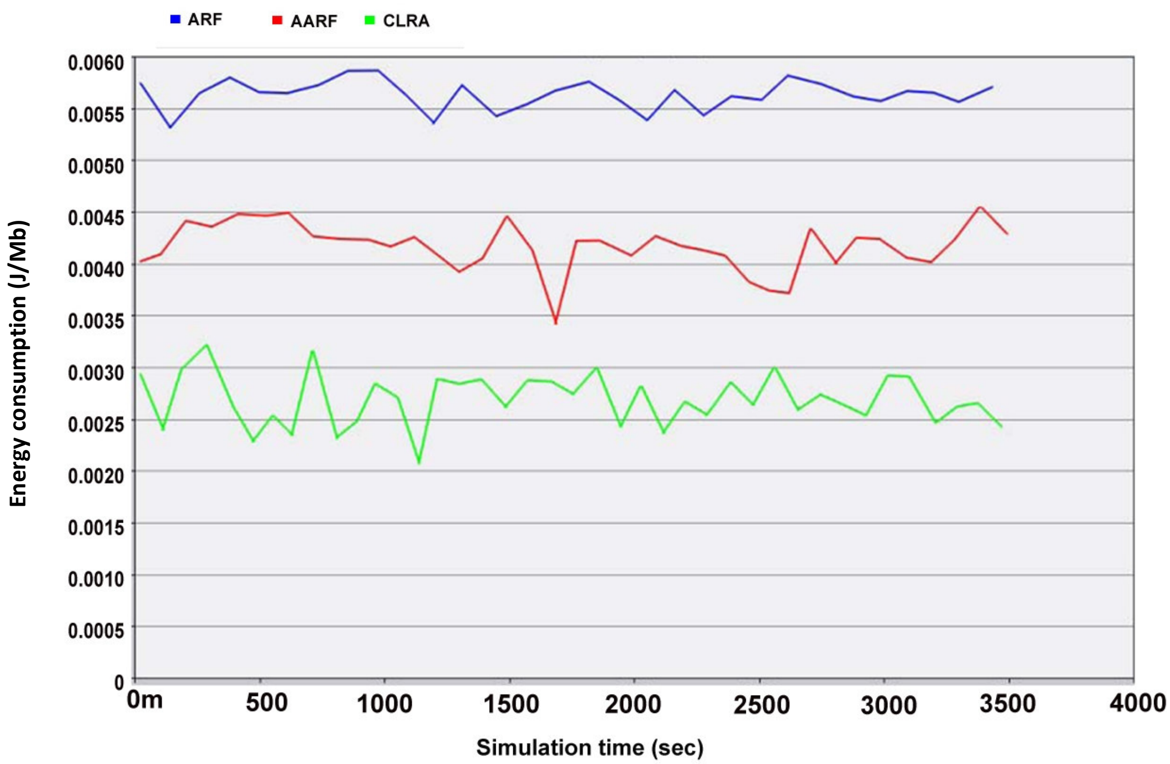

Figure 4-15: Energy consumption comparison in three scenarios using ARF, AARF and CLRA.

4.2.8 Conclusions

The motivation behind proposing CLRA as an alternative rate-adaptation solution is to determine an instantaneous value of the higher transmission rate which a rate adaptation scheme should attain under the existing conditions. The motivation is partly based on the fact that CLRA should avoid the chances of encountering frame-losses (and thus retransmissions) to the best; and partly the motivation comes from the fact that even in ideal channel quality conditions, it is not a requirement to transmit at the highest possible transmission rate.

Combining the two reasons, we came up with a solution to determine a runtime requirement of application's generated traffic, then determine, the underlying medium access constraints (and thus estimate the contention delays), and then select such a value which is most suitable and at the same time minimizing the chances of retransmissions.

The proposed scheme is self-adjusting according to the medium contention levels. Therefore, a station whose application generated traffic is very low would not be able to penalize the rest of the stations by operating at a lower rate. In this situation, the medium contention becomes higher and thus the CLRA limiting values are also incremented.

CLRA uses a window RTS/CTS exchange for differentiating *frame collisions* from *frame corruption*. The mechanism is effective in determining the presence of hidden nodes and then adjusting the rate-adaptation scheme. However, the use of RTS/CTS is a communication overhead. The proposed frame-loss-differentiation mechanism uses the RTS/CTS procedure effectively to minimize the overhead.

4.3 Performance tests and comparative analysis of MutFed

This section discusses the comparison of MutFed with ARF and AARF in various test conditions. The main goal is to:

- Compare the responsiveness of these rate-adaptation schemes to variations in channel conditions.
- Compare the inherent overhead in terms of retransmissions and the resultant delay during long-term channel variations.
- Analyze the overall efficiency of these rate-adaptation schemes by using throughput at the receiver side as a measure.

In order to compare ARF, AARF and MutFed, separate simulation models for the rate-adaptation schemes were created at the MAC-layer. The default model for modelling MAC-layer functionality in OPNET simulator is called wlan_dispatch. The wlan_dispatch is a parent process with two child processes i.e. the wlan_mac for modeling the 802.11 standard specification and wlan_hcf which models the QoS extension as specified in IEEE 802.11e.We modified the wlan_mac process, by writing our code at appropriate places in various states and functions which are invoked from those states at various instances.

In the simulation scenario, a transmitter station is placed very close to a receiver where it stays for one minute; let's call it position 'A'. The transmitting station then moves further to position 'B', 'C', 'D', 'E' and finally to 'F', staying at each of these positions for one minute. From 'F', the transmitter moves back to its original location which is point 'A'. To simulate sudden variation in channel quality, the transmitter's movement from position 'F' back to the initial position 'A' is done at considerably higher speed. The type of traffic exchanged between the transmitter and receiver is of variable-bit-rate (VBR) and the frame sizes were also variable.

The analysis of simulation scenarios shows that both ARF and AARF try to increment the transmission rate after a predefined number of successful frame transmissions. Using predefined thresholds for rate-up attempts have serious implication on the communication efficiency especially when the channel quality deterioration lasts for a longer time. For instance, when the transmitter moves away, the received signal power at the receiver gradually drops down. ARF and AARF forced to reduce the transmission rate after two consecutive retransmissions. However, the channel quality sensing mechanism in ARF and AARF is based on heuristically designed thresholds. Therefore, while operating at the lower rate and while the transmitter station is still on the move (moving away from the receiver, in this case), if the rate-up thresholds are reached, both of these rate-adaptation scheme would send next frame at a higher rate. Such rate-up attempts, in similar scenarios, are completely *blind-channel sensing attempts,* causing guaranteed frame failures. If the transmitter stops and stays for a longer time at either of the positions, ‘C’, ‘D’, ‘E’, or ‘F’, such unnecessary retransmissions would be periodically seen both in the case of ARF and AARF. A similar behaviour will be shown by SampleRate.

The frequency of rate-up attempts in the case of ARF is higher than AARF. So, when the transmitter moved from A-B-C-D, both of these rate-adaptation schemes lowered the transmission rates one step at a time. However, with every 10 frames (in case of ARF) a rate-up attempt was made; this tendency of ARF and AARF caused considerable number of retransmissions. In case of MutFed, the rate-up action is recommended by the receiver based on the average SNR values i.e. it does not make blind rate-up attempts to check improvements of channel quality, which reduces the number of retransmissions.

Figure 4-16 shows the comparison of the number of retransmission in the same simulation environment by ARF, AARF and MutFed. Here the collection mode for recording the number of retransmissions is ‘bucket’ with a size of 1000 values, and each value which is plotted in the Figure 4-16 is the sum of the bucket-size. In simple words, a value of 100 in Figure 4-16 means 10% retransmissions. For ARF, the average retransmissions are approximately 21% and for AARF the number is reduced to approximately 10%, while for MutFed the number of retransmissions is approximately 3.5%.

Every communication system pays a high price for retransmissions. In the case of IEEE 802.11, a retransmission forces the MAC to increment the CW, resulting in relatively higher backoff values, which implies that stations which encounter retransmission are forced to wait for a longer time before accessing the medium. Likewise, a retransmission is interpreted by ARF and AARF as a sign of ‘bad’ channel quality and with two consecutive retransmissions they reduce the transmission rates. Therefore, because of the lack of actual reasons of channel quality deterioration and the duration of channel quality variations, ARF and AARF use deterministic thresholds to perform rate-ups as a method to check the CSI.

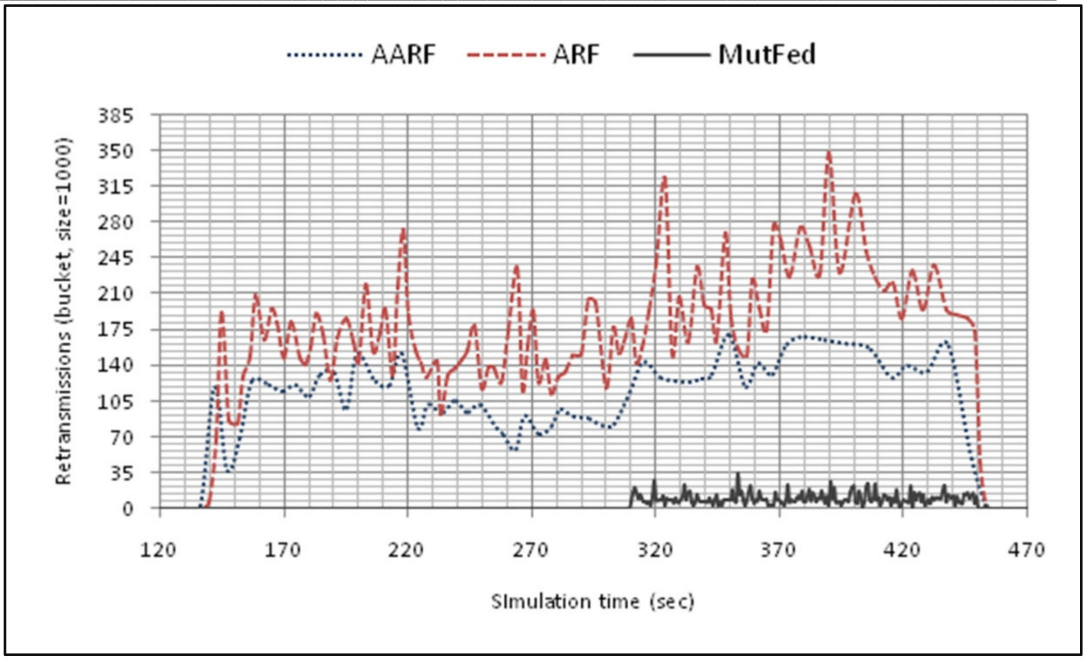

Figure 4-16: Comparison of frame retransmissions in a similar scenario using ARF, AARF and MutFed.

This behaviour (of ARF and AARF) causes undue (and reasonably avoidable) frame losses and inefficiency in communication; this not only affects the performance of a single station but affects the overall BSS.

The end-to-end delay between the transmitter and a receiver during frame transmission is calculated by including the backoff time, the number of retransmissions and the propagation time. Normally, the end-to-end delay can be ideally calculated at the receiver, by subtracting the frame transmission time from the current time. However, in this case, in order to include the effect of increasing backoff values as a measure of medium access delay and the retransmissions, the end-to-end delay is calculated at the sender side. In the IEEE 802.11 standard, when a sender sends a frame, it waits for ACK_timeout duration. The ACK_timeout duration is given by: SIFS time + SlotTime + PHY-RX-START-Delay.

The backoff counter is incremented every time, when there is a retransmission. This is done in the *forced-state* of 'PERFORM BACKOFF' as shown in Figure 4-1. So, for instance, the delay measure gives the delay experienced while waiting for the medium (access delay), this includes the backoff time, the propagation time (which depends on the PPDU size and the transmission rate). Therefore, using this measure, a frame which is retransmitted would have higher delay because of the higher medium access delay and (possibly) higher propagation time. The comparison between ARF, AARF and MutFed, in terms of time-average delay is given in Figure 4-17.

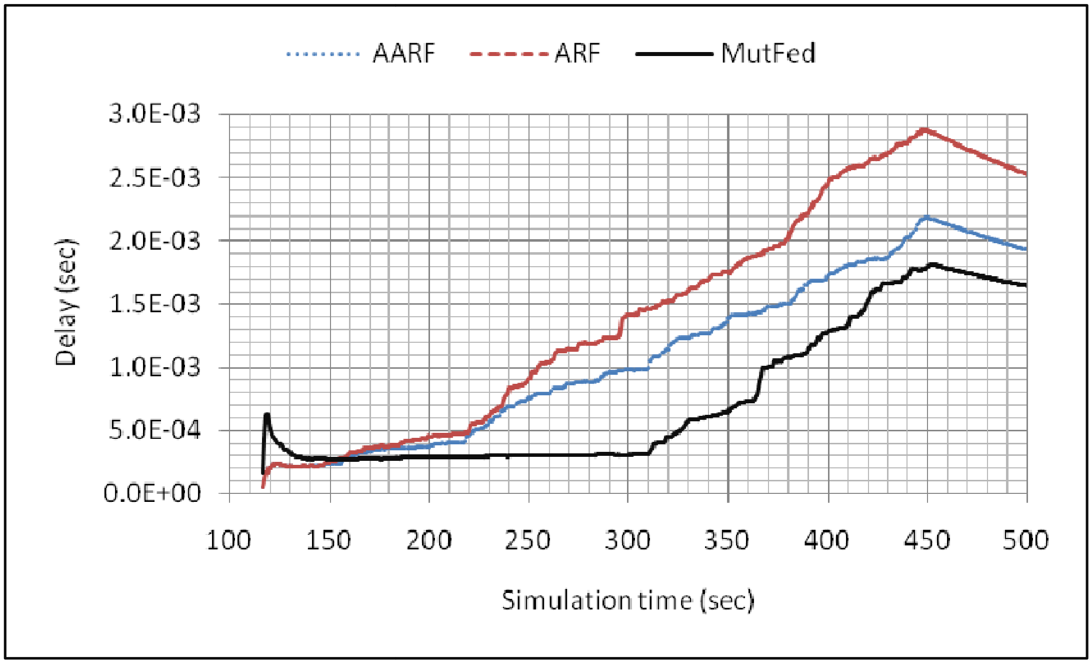

Figure 4-17: Time-average delay comparison of ARF, AARF and MutFed.

A comparison of throughput at the receiver (in terms of received bits per second) shows that ARF performs poorly when compared to AARF. ARF performs a rate-up after a fixed interval, but AARF holds back for longer time at lower rates if probe frame fails. Therefore, at the time, when the distance between the transmitter and receiver is maximum, because of the undue (and possibly avoidable) retransmissions AARF and ARF stay at lower rate, with the former staying for comparatively longer. While in case of MutFed there are no unnecessary rate-up attempts which avoids excessive retransmissions and the transmitter operates at relatively higher rates during the same channel conditions.

The comparison of throughput of ARF, AARF and MutFed is given in Figure 4-18. It is evident that throughput *dips* in case of ARF are higher and more frequent than AARF and MutFed. Figure 4-18 shows that ARF's throughput is the lowest when channel quality is at its worst state i.e. when the transmitter is at the furthest distance). However, because of its inherently higher frequency of rate-up attempts, it responds to channel quality improvements much faster than AARF; this can be noticed from Figure 4-18 at the simulation time of 420 until 430 seconds.

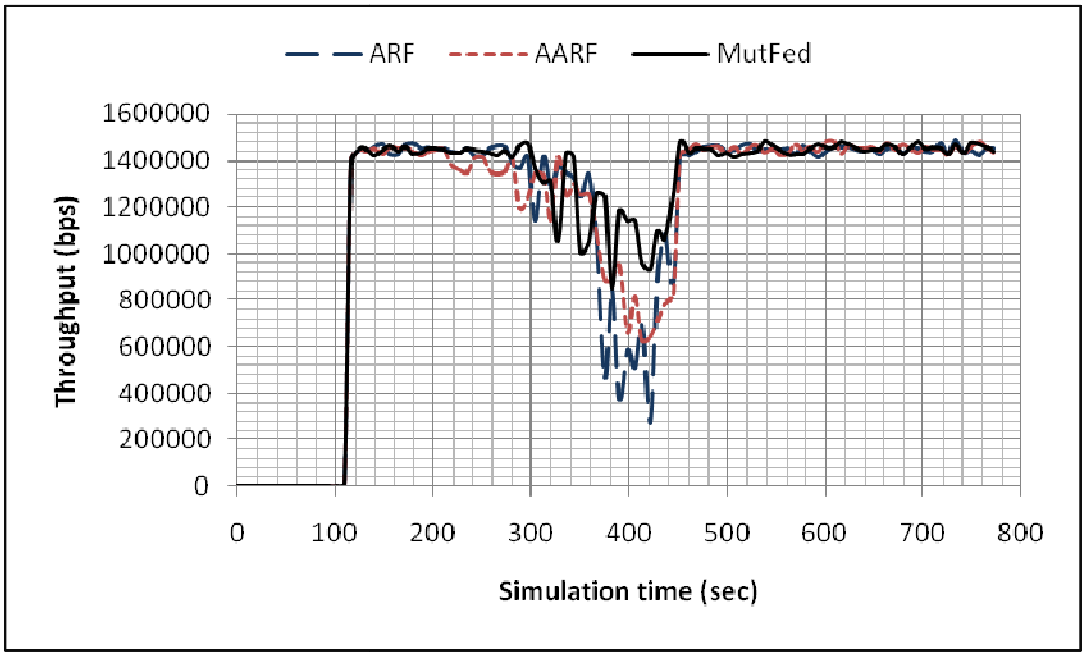

Figure 4-18: Throughput comparison of ARF, AARF and MutFed.

It seems logical to know about the maximum achievable throughput in the simulated scenario which is under discussion. To find the maximum achievable throughput, six different simulation scenarios were configured; no rate-adaptation scheme was used in any of the six scenarios. The transmitter and receiver were both configured to use a fix transmission rate of 1, 2, 11, 24, 36 and 54 Mbps. This would enable us to answer the question that:

Which transmission rate would perform best (in terms of throughput) at *what* time?

The throughput recorded at the receiver for scenarios where the transmission rate is 54, 36 and 24 Mbps, is shown in Figure 4-19. It is obvious from Figure 4-19 that transmissions at higher rates result in transmission disruption with the receiver showing zero throughput as the distance between the transmitter and receiver pair increases. Figure 4-20 shows throughput for 11, 2 and 1 Mbps scenarios. A joint comparison of throughput is shown in Figure 4-21.

It is evident that the performance of MutFed is clearly better than ARF and AARF. When the channel quality variation starts initially because of the transmitter's movement, MutFed detects this change from the received signal power and as a result the transmitter quickly drops down the transmission rate as suggested by the receiver. However, in the case of ARF and AARF, there are periodic rate-up attempts. While the transmitter is on the move, there are short intervals when transmission at two adjacent transmission rates is possible. Such intervals are at times used by ARF and AARF to transmit at higher rates.

For instance, suppose that the transmitter is on the move and currently transmitting at 36 Mbps. If there are frame failures, all rate-adaptation would lower the rate. In case of MutFed, the receiver would approve the rate-down decision of the transmitter as it is very much likely that the received power of the latest frames would be lower than that it has for the current-window. As a result, after lowering the transmission rate to 24 Mbps, MutFed would stop making unnecessary rate-up attempts at 36 Mbps, unless the receiver recommends it. On the other hand, both ARF and AARF would periodically transmit at 36 Mbps. For very short intervals, transmission may likely be successful at 36 Mbps. Therefore, in such short intervals, the throughput of ARF and AARF would be higher than MutFed. This is trait can be seen in the Figure 4-19 between the interval from 310-370 seconds.

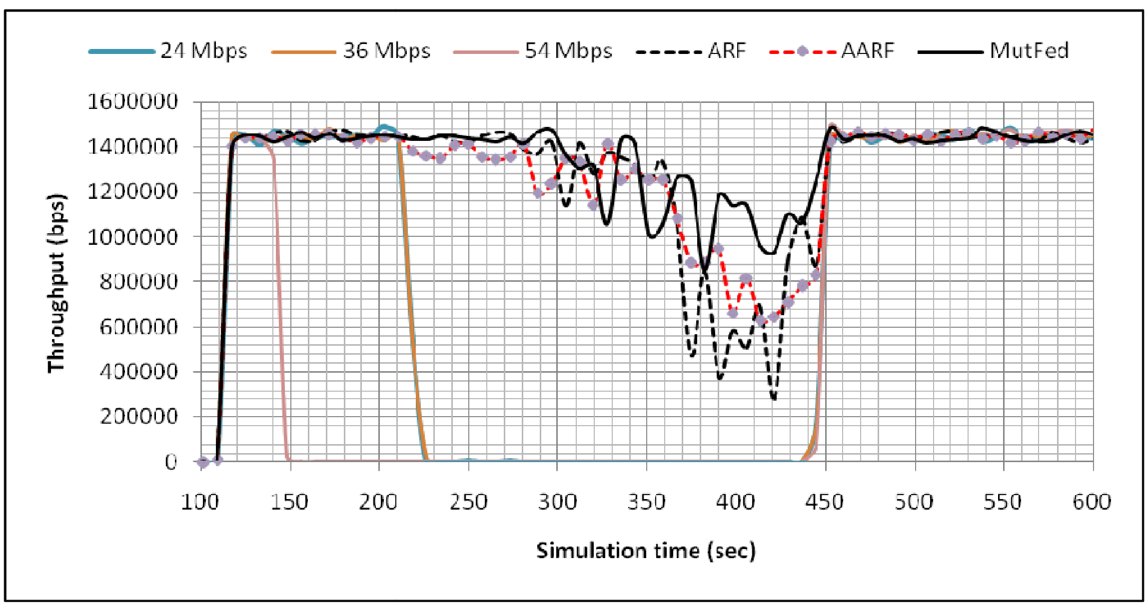

Figure 4-19: Throughput analysis at various transmission rates (24, 36 and 54 Mbps) and ARF, AARF and MutFed in the simulation scenario.

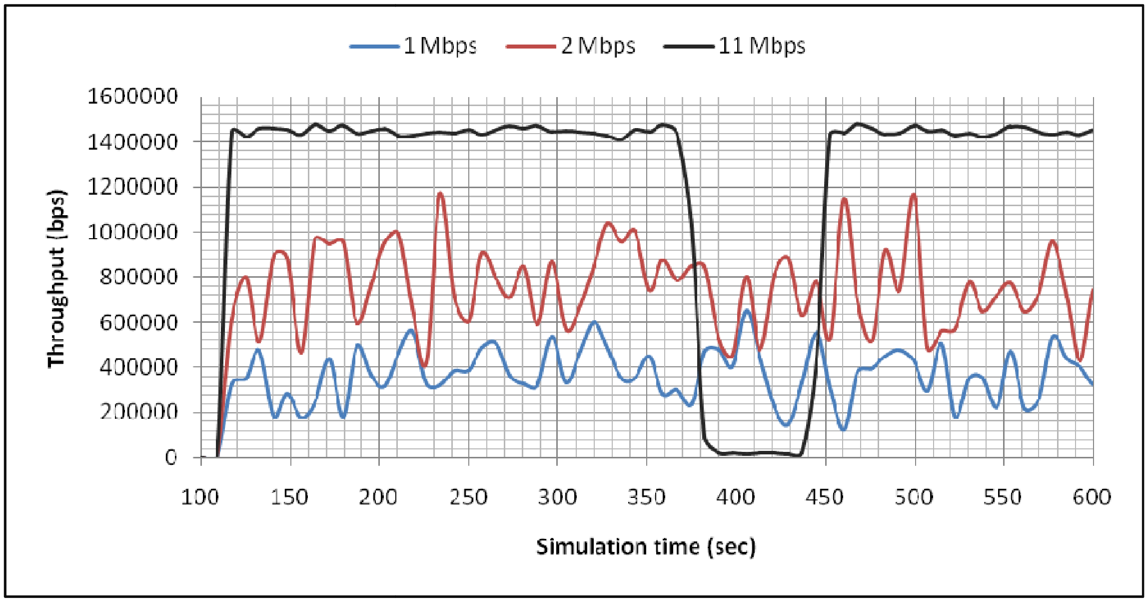

Figure 4-20: Throughput analysis at various transmission rates (1, 2 and 11 Mbps) and ARF, AARF and MutFed in the simulation scenario.

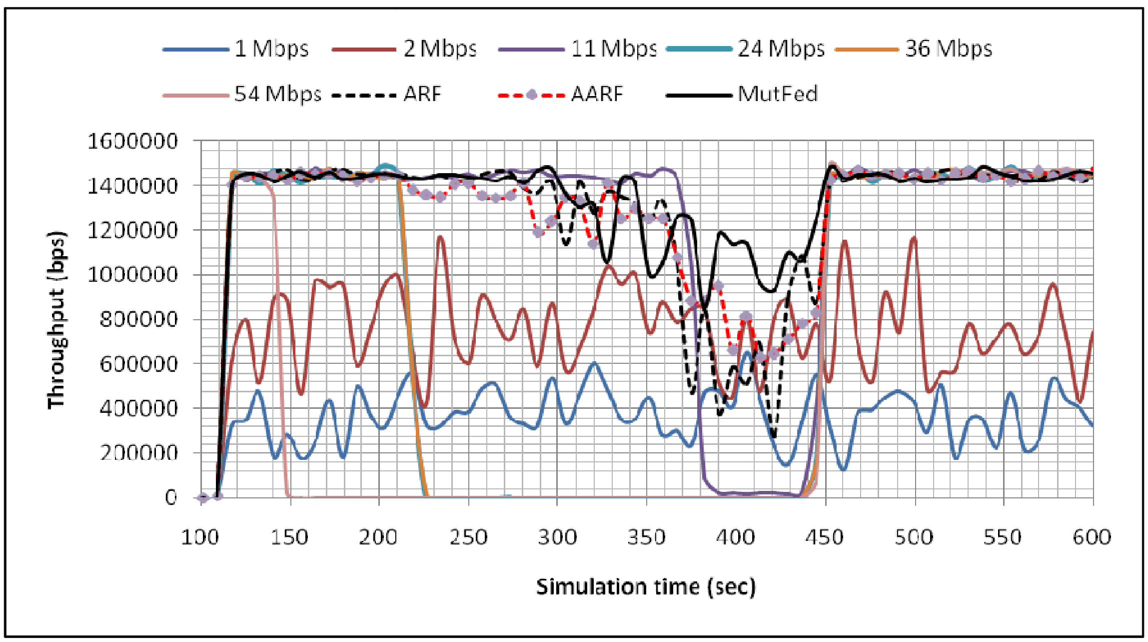

Figure 4-21: Throughput analysis at various transmission rates and ARF, AARF and MutFed in the simulation scenario.

4.3.1 Frame-loss differentiation mechanism: *comparison*

A comparison of the frame-loss differentiation mechanisms used in CARA [5], CLRA and MutFed is shown in Table 4-1. Here in this table, R1, R2 and R3 represents various transmission rates which are selected by the rate-adaptation schemes where R1>R2>R3 and so on. The notations 'L' and 'NL' represent frame transmission status and stands for 'Lost frame' and 'Not-lost frame' respectively. Here, we use the assumption that channel quality remains unchanged and frames are lost only because of simultaneous transmission by hidden terminals.

There are two cases; in case-A the frame losses in the first frame window do not occur consecutively. In this case, the performance of MutFed is evidently better than the two schemes. MutFed sent 100% frames at R1, with no overhead

In test case-B, the first frame window has consecutive L-frames, which forces all the rate adaptation schemes to reduce the transmission rates. In this case, MutFed performs better with only 6.25% frames sent at a lower rate. That is the time, when the receiver detects that lower rate selection is a wrong decision, as we mentioned earlier, the channel quality remains the same during this duration.

Therefore, MutFed's loss differentiation mechanism is not only highly responsive; it also avoids unnecessary reduction of transmission rates.

Table 4-1: Comparison of the frame-loss differentiation mechanisms used by MutFed and CLRA. Here in this table, R1, R2 and R3 represents various transmission rates which are selected by the rate-adaptation schemes where R1>R2>R3 and so on. The notations 'L' and 'NL' represent frame transmission status and stands for 'Lost frame' and 'Not-lost frame' respectively.

A																
	frame x	frame x+1	frame x+2	frame x+3	frame x+4	frame x+5	frame x+6	frame x+7	frame x+8	frame x+9	frame x+1	frame x+1	frame x+1	frame x+1	frame x+1	frame x+1
Rate adaptation scheme	L	NL	L	NL	NL	NL	NL	NL	L	NL	L	NL	L	NL	NL	NL
CLRA	R1	R1	R1	R1	RTS R1	RTS R1	RTS R1	RTS R1	R1	R1	R1	R1	RTS R1	RTS R1	RTS R1	RTS R1
MutFed	R1	R2	R1	R2	R1	R1	R1	R1	R1	R2	R1	R2	R1	R2	R1	R1
	CLRA	sent 100% frames using R1 with 50% transmission overhead because of the use of RTS/CTS procedure.														
	MutFed	sent 68% frames using R1, and 31.25% frames using R2, with 0% transmission overhead.														
B																
	frame x	frame x+1	frame x+2	frame x+3	frame x+4	frame x+5	frame x+6	frame x+7	frame x+8	frame x+9	frame x+1	frame x+1	frame x+1	frame x+1	frame x+1	frame x+1
Rate adaptation scheme	L	L	L	L	NL	NL	NL	NL	L	NL	L	NL	L	NL	NL	NL
CLRA	R1	R1	R2	R2	RTS R3	RTS R3	RTS R3	RTS R3	R2	R2	R2	R2	RTS R1	RTS R1	RTS R1	RTS R1
MutFed	R1	R2	R3	R4	R1	R1	R1	R1	R1	R2	R1	R2	R1	R2	R1	R1
	CLRA	sent 37.5% frames using R1, 37.5% at R2 and 25% at R3 with an overall 50% transmission overhead because of the use of RTS/CTS procedure.														
	MutFed	sent 62% frames using R1, and 25% frames using R2, 6.25% of frames at R3 and R4 with 0% transmission overhead.														

4.3.2 Conclusions

MutFed is a receiver side-SNR based rate-adaptation scheme which uses receiver's feedback while selecting a transmission rate for future transmissions. Unlike the previous receiver-side (closed-loop) rate adaptation schemes MutFed do not require modification of the standard frames.

In section-4.3 we analyzed the MutFed rate-adaptation scheme with the help of simulation tests. Comparison with ARF and AARF were shown in various test scenarios. It is concluded from the comparative analysis that MutFed is highly responsive when compared to statistics based rate-adaptation schemes and converge quickly to the most suitable transmission rate under prevailing channel quality. MutFed does not rely on frame deliver statistics and thus the rate-up and rate-down attempts are more realistic and guided by the actual SNR. MutFed also introduces a novel mechanism of delivering the feedback information from the receiver back to the transmitter. The rate-selection feedback which is suggested by a receiver is delivered without any modification to the standard frames as has been suggested in the previous research literature.

MutFed uses a novel and extremely efficient mechanism for differentiation of frame losses. The frame loss differentiation is performed without the use of RTS/CTS exchange, thus avoiding communication overhead. Moreover, loss differentiation is achieved just with the help of three frames. The loss-differentiation mechanism enables the rate-adaptation process to work in an intelligent manner.

4.4 References:

[1]. OPNET Modeler, OPNET Technologies Incorporation. www.opnet.com

[2]. Kamerman and L. Monteban, "WaveLAN®-II: a high-performance wireless LAN for the unlicensed band," Bell Labs Technical Journal, vol. 10, pp. 119, 1997.

[3]. M. Lacage, M. H. Manshaei and T. Turletti, "IEEE 802.11 rate adaptation: a practical approach," Proceedings of the 7th ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems, pp. 126-134, 2004.

[4]. J.C. Bicket, Bit-rate selection in wireless networks, M.S. thesis, MIT, Feb 2005.

[5]. J. Kim, S. Kim, S. Choi and D. Qiao, "CARA: Collision-Aware Rate Adaptation for IEEE 802.11 WLANs," 25th IEEE International Conference on Computer Communications, INFOCOM 2006. Proceedings 1-11, 2006.

Chapter 5

Medium Aware Distance Vector (MADV) Routing Protocol

5.1 Introduction

Mobile ad-hoc networks (MANETs) have received an exponentially growing attention from the research community right from its inception in the early 1980's through the Defence Advanced Research Projects Agency's (DARPA) projects [1-2]. In contrast to wired networks, MANETs are characterized by dynamic topological changes due to station mobility; such a trait of MANET makes the task of establishing routes a very challenging process. Early efforts for MANET routing protocols were inspired by the methodologies used in routing protocols for wired networks. Such trends lead to several MANET routing protocols, where route selection process essentially tried to select routes with minimum distance between the source and destination stations and so the 'number of hops' between the source and destination became a default routing metric.

However, the characteristics of wireless networks are significantly different from their wired predecessors. In case of the wireless networks, link quality stands out as a very important parameter in determining the communication quality. Link quality is a generic term which encompasses several quality metrics and these parameters in turn determine the quality of communication over a certain link such as the frame error rate, the transmission rate which is used by a transmitter under prevailing signal to noise ratio (SNR) while maintaining an acceptable frame error rate, the number of allowable retransmissions while meeting a certain efficiency threshold etc. Link quality in general is variable and depends on a mobile station's mobility patterns, interference from external sources and so on. Therefore, it is important to consider link quality as a significant factor while determining routes between source and destination stations.

Wireless communication stations use a shared frequency band where access regulation to a particular frequency band is coordinated with the help of medium access control (MAC) protocols. The MAC protocol ensures that only one station accesses the medium at one point in time to avoid collisions. In situations of higher station density within the transmission range, stations are put on hold for longer. In case of the IEEE 802.11 standard compliant stations, they may have to freeze their backoff counters a number of times before they find an opportunity to transmit on the medium. Transmission errors and station density would account for higher medium occupancy by neighbouring stations and as a result every station would experience higher contention delay for accessing the medium. Likewise, stations using lower transmission rate can penalize neighbouring stations by causing higher medium contention. Therefore, a station may be able to transmit at a higher transmission rate; however, because of medium occupancy the medium access delay would be higher. In such situation, a route having higher throughput links would encounter higher end-to-end delay. Therefore, a routing protocol for MANET should take into account such MAC specific parameters and determine their implication on the performance of end-to-end communication.

The end-to-end contention delay can potentially increase with increase in the number of stations along the route. Moreover, a routing protocol which selects a route by including links with higher transmission rates would generally imply that the route has a higher number of intermediate stations between a source and destination. Higher number of intermediate stations means relatively higher contention delay along the route. In addition, in a multi-rate multi-hop network, the end-to-end throughput is capped by a station using a relatively lower transmission rate along the route. If lower transmission rate links follow higher transmission rate links, frames would pileup in the transmission queues at a station transmitting at a lower rate and as a result there would be higher queuing delay [3]. The same is true when packets, on its way through the multi-hop route, land in a highly congested area. Avoiding, congested (having higher medium access contention delays) links and constructing a route with relatively less contention and queuing delays necessitates incorporation of MAC and PHY specific information in the routing protocol.

Link quality parameters (e.g. transmission rate, frame error rate, signal to noise ratio) are rooted at the physical layer of the OSI architecture. Likewise, monitoring the medium contention levels, frame prioritization and queuing delays are dealt at the medium access control (MAC) sub-layer. The need of inclusion of PHY and MAC level information in the routing protocol as mentioned above makes such a design a classical case of cross-layer architecture. Cross-layer information exchange has been presented as an effective and inevitable solution in many cases [4-8].

This chapter gives detail of our proposed routing protocol called Medium Aware Distance Vector [8](MADV). MADV formulates a routing metric which combines variations in the transmission rates, contentions levels and queuing delay at every link. A significant addition of MADV is the incorporation of the queue prioritization information which is introduced by the IEEE 802.11e MAC specification. The IEEE 802.11e introduces four different MAC level queues for each of the four access categories. The standard specifies maintenance of service priority within the queues; which implies that frames from a higher priority queue would be serviced more frequently than frames belonging to lower priority queues. Such an enhancement at the MAC sub-layer introduces uneven queuing delays. Conventional/existing routing protocols are unaware of such queue prioritization and as a result these factors are not considered which result in severe performance deterioration for frames belonging to lower priority queues. It is important to consider hop-by-hop (or a link) MAC level information while constructing an end-to-end routing strategy for a multi hop wireless network. In MADV the routing table structure is modified from the existing routing protocol's approaches by introducing per destination-per AC entry; which essentially means that a source station can have different route entries based on the ACs for the same destination. The advantage of ACs-specific route entries is that MAC-level queue prioritization can be effectively taken into account for the whole route, and the route selection is performed based on the lowest metric value per AC from the available routes to a destination.

[8] 'Distance Vector' is a term used for classifying routing protocols according to the way routing protocols operate. The term 'distance' refers to cost of reaching a destination on a particular interface (which is referred to as the 'vector') of a source station. The cost is the routing metric's value. Therefore, in a distance vector routing protocol, every station constructs a list of destinations that it can reach on a particular vector (interface) and indicates the cost of reaching there (i.e. the metric value).

The key contributions of this chapter are as follows:

1. Significance of variation of the transmission rate, the use of a rate-adaptation scheme at the MAC layer and its effects on the performance of routing protocols are investigated in detail and key issues specific to routing protocols are highlighted.

2. The significance of medium occupancy, PHY and MAC level factor contributing to this phenomenon and the effects of medium contention on the performance of route selection.

3. An in depth analysis of the IEEE 802.11e inter-queue access priority, the resultant queue delay variations and their effects on routing strategy. These three factors lead to the formulation of the MADV routing metric.

4. Design of a route selection metric based on the per-AC-per-destination, which is completely different too from the existing/conventional routing protocols.

5. Formulation cross-layer information exchange frame work for communication MAC and PHY specific parameters for inclusion in the routing protocol route metric.

6. A novel route selection mechanism for MANETs, which incorporates the underlying MAC and PHY specific parameters and the overall route selection is based on the actual state of the medium.

This chapter is divided into the following section: section-5.2 discusses in detail various classes of MANET routing protocols, section-5.3 gives an overview of the MAC and PHY constraints in communication and various factors which arise from these constraints and the effects of these factors on routing protocols, section-5.4 presents an analysis of various research efforts focusing on enhancements to routing protocols in connection with incorporation of medium awareness, section-5.5 gives detail of the MADV routing protocol, section-5.6 explains the comparative analysis of the MADV routing protocol in various simulation scenarios, and finally section-5.7 concludes the chapter.

5.2 An Introduction to MANET routing protocols

MANET routing protocols are classified into various categories. The classification depends on the mechanism of *route discovery*, *maintenance* and *dissemination* of routing information across the MANET.

5.2.1 Proactive Routing protocols

As the name suggests, proactive routing protocols maintains routing information about the network in a proactive manner. Mobile stations which use proactive routing protocols exchange their routing tables with neighbours about all discovered routes to all destinations. In this way, routing information is propagated across the network for all connected stations in a network. Therefore, (after a certain time) a source station should immediately find an entry to a destination station using proactive routing protocols. Routing protocols belonging to this class are also called '*table driven*' routing protocols.

Proactive routing protocols, maintain consistent routing information and in the event of any route changes, the updates are then sequentially propagated across the network.

Proactive routing protocols inherently consider seldom changes in the underlying network connectivity and topological changes; an assumption which is rooted in the routing for *wired* networks. Due to longer life of topology in case of wired networks, after the initial convergence of the routing tables (when almost all destinations are accessible), there is no frequent demand for route updates and thus no extra overhead which is caused by control packets for propagating the route updates.

However, in the case of MANETs, topological changes are variable both in terms of frequency, and in terms of the number of stations. Therefore, the mechanism of maintaining routes to all stations (even if they are never been required) by every station in MANET, comes at a high price in terms of route update information.

Some of the highly cited proactive routing protocols are briefly discussed below.

5.2.1.1 Destination-Sequenced Distance Vector (DSDV).

DSDV [9], is one of the highly cited proactive routing protocols. In DSDV, every station in MANET, maintains its own routing table, having entries for every possible destination and the number of hops to each destination. The route-table entry for every destination is tagged with a sequence number by the destination station. In DSDV, stations periodically exchange information in their routing tables with their neighbouring stations. In this way, changes in the network topology can be propagated across all the MANET. When a mobile station receives new information (a route update), it compares that information with the entries in its routing tables. If the latest received update has a higher (new) sequence number for a specific destination station, the mobile station updates that entry. It simply discards the update if the sequence number is older than the sequence number of already existent routing entry. In case if the sequence numbers are same, but the number of hops to a specific destination in the latest update is less than the already existent entry, it updates the routing table.

Newly recorded routes are scheduled for immediate advertisement to the neighbouring stations, while route entries with better route metric are scheduled for time which depends on average route settling time. Therefore, route updates are *event driven* or *time-driven*. The update mechanism in DSDV is in the form of either a *full dump*, consisting of several packets, where in full dump the whole routing entries are sent to the neighbours, or the DSDV may use single packet update, called incremental updates, for only those routes where the hop count changed.

5.2.1.2 Optimized Link State Routing Protocol (OLSR)

OLSR [10], is a table-driven, proactive routing protocol for MANETs. In OLSR, every station selects a set of its neighbours as “multipoint relays” (MPRs); such MPRs are responsible for forwarding routing related control traffic. The use of MPRs potentially reduce the amount of traffic which is generated as a result of the flooding of control messages. Broadcasts are controlled in a way that when a station sends a broadcast message, all stations which are in its MPR set would receive such a broadcast, but would not further forward it. An MPR maintains a list of stations which has selected it as an MPR using periodic HELLO messages.

Each station maintains routing table with entries for all destinations. Each entry has a destination address for the destination station, the next hop neighbour address and the number of hops to the destination. Route entries are updated in the event of change in the neighbour connectivity, 2-hop neighbour set, etc.

The use of MPRs relieves the MPR-selectors stations from the responsibility of route updates. The MPR stations declare the link state information of its selector stations in periodic control messages. Such control messages indicates that all stations mentioned by an MPR are accessible through that specific MPR.

5.2.1.3 Topology Dissemination Based on Reverse-Path Forwarding (TBRPF)

TBRPF [11], is another proactive routing protocol which provides routes in a MANET using the shortest path to the destination. Each station in the TBRPF maintains a source tree, which provides shortest paths to all reachable stations. Each station computes and updates its source tree based on partial topology information stored in its topology table, using a modification of Dijkstra's algorithm. Stations only exchange a part of its source tree information with its neighbouring stations for minimizing the overhead. This part of the source tree which is exchanged with neighbours is called reported sub-tree. Each station sends the reported sub-tree using periodic update interval and changes in the source tree are reported using a more frequent differential update interval. The periodic updates inform newly joining neighbours about the reported sub-tree and differential updates inform all the stations which are affected by such changes in a quicker way.

5.2.1.4 Fisheye state routing (FSR)

FSR [12], uses the concept of fish eye view, where the nearer objects are seen with more clarity and distant objects are relatively unclear. In the FSR based routing, every station maintains a view of the network topology, by maintaining topology tables based on the information received from the neighbouring stations. Stations send periodic updates to other stations. In order to reduce the update propagation across the whole network the frequency of updates for neighbouring stations (closer to a station) is higher and for further stations, the frequency is lower. This is accomplished by using different update intervals for various entries in the routing table.

For every station the network is divided into near and far zones and its routing updates propagation frequency is limited by this distance. The use of variable frequency for updates to distant stations enhances the scalability of FSR when compared with conventional proactive routing protocols.

5.2.2 Reactive routing protocols

In contrast with proactive routing protocols, this class of routing protocols search for a route to a particular destination in MANET only on the basis of when a demand for that destination arises. Routing protocols belonging to this class are also called 'On-demand' routing protocols.

The motivation behind using an on-demand path search is linked with the growing routing related control traffic in the proactive routing protocols. Such control traffic overhead increases as the network size grows and is also dependent on the stations' mobility patterns which result in different topological changes. All such changes, needs updates across the network. In the case of on-demand routing protocols, no route entries are maintained for destinations which are not needed, therefore, in the event of route changes to destinations-of-no-concern, all stations do not need awareness of the changes. By doing so the routing tables' size is minimized, minimizing the route maintenance and related periodic updates.

Some of the highly cited reactive routing protocols are briefly mentioned below:

5.2.2.1 Ad-Hoc On-Demand Distance Vector (AODV)

AODV [13], is an on-demand routing protocol. The route discovery starts, when a source station initiates communication with a destination station for which the source station does not have a valid route entry. The route discovery process consists of broadcasts of Route Request (RREQ) messages. Such broadcasts originate from the source and propagated towards the destination. Intermediate stations process the RREQ messages and based on their routing entries in their local tables, process the RREQ messages accordingly.

An originator station uses a *destination sequence* number and *originator sequence* number in the RREQ message. The destination sequence number shows the previous sequence number known to the current station about the route to the specific destination. In case of no previous sequence number information about the destination station, the unknown sequence number FLAG is set. The originator sequence number shows the sequence number of the originator station; this number is incremented from its previous value before using the in the RREQ. A hop count field which indicates the number of hops between the source and destination station is set to zero by the originator station.

The RREQ message can be processed by an intermediate station which has information about the destination station. However, in many cases, the route between two stations is used for bi-directional traffic flow. Therefore, when the RREQ is processed by an intermediate station, the destination may be informed about the route towards the source station by setting the 'G' flag, for generating the Gratuitous RREP back to the destination of a RREQ.

Route discovery using broadcast RREQ can span the whole network. In order to control the dissemination of such RREQ messages across the whole network, the originator station first sets the value of time to live (TTL) field in the IP header to TTL_START and waits for receiving a corresponding RREP within RING_TRAVERSAL_TIME. If a RREQ is not received within the timeout period, the TTL duration is incremented by TTL_INCREMENT. This process continues till the TTL reaches TTL_THRESHOLD, after which the TTL is set to NET_DIAMETER.

A station receiving a RREQ message processes this message in various ways, depending on the information in its routing table. Initially, it increments the hop count field value by one to account for the one hop and then checks if it has reverse route information back to the originator of the request; such route entry is updated using the originator sequence number as the destination sequence number for the originator's address in the current stations routing table. If the receiving station does not have information about the destination station, it cannot generate RREP, therefore, it creates a RREQ field, by decrementing the TTL value, and setting the destination sequence number to the highest sequence number known to this station for that specific destination. This station then rebroadcasts the RREQ further.

RREP is generated by a station if the station is either the required destination or it has a valid route entry for the destination station. A destination station updates its sequence number if the sequence number for this destination in the received RREQ is greater than its current sequence number. It resets the hop count field to zero. Similarly, an intermediate station also updates its sequence number for the destination accordingly and places the hop count value that it has for the destination in the RREP message.

A station receiving the RREP messages update their routing entries for the destination station in terms of the destination sequence number and smaller number of hops.

In order to remain listed as an active station, every station using the AODV periodically broadcasts HELLO messages. Every HELLO_INTERVAL a station checks whether it has sent broadcast within the last HELLO_INTERVAL. The HELLO message is broadcasted with a TTL of 1, mentioning its own address and the latest sequence number for the current station. The TTL value of 1 restricts the dissemination of HELLO message only to one hop neighbourhood and refreshes the neighbour connectivity every HELLO_INTERVAL.

5.2.2.2 Dynamic Source Routing (DSR)

DSR [14], is another variant of reactive routing protocols for multi-hop wireless networks using a rather different mechanism of *source routing.* In source routing, each data packet which is destined for a station in the network, carries a complete ordered list of addresses of intermediate stations along the route between the source and destination stations. In such form of routing (source routing), the sender station controls and selects the routes used for its own packets. Including the source selected route information in the header of every data packet enables intermediate stations to overhear the routing information to various stations in the network; such overheard information can then be used for their own routing purposes.

A station initiating a data exchange with a station for which it does not have a valid route entry, starts a route discovery process. In the route discovery phase, it sends a broadcast and includes its own address; such message contains a unique identification number to identify the requesting station and the destination. When an intermediate station receives such a request, it either sends a RREP if it has a valid route to the destination by inserting the complete route records in the RREP. Otherwise, it inserts its own address and rebroadcasts the message to its neighbouring stations.

A destination station can simply reverse the route record received in the RREQ message and use it as source route from itself to the initiator station.

DSR uses a mechanism of *packet salvaging*; under this mechanism when an intermediate station determines that the next hop as indicated in the source route is broken, it searches in its local cache for other possible routes to the same destination and instead of discarding the packet, it updates the route information to include the new route to the destination.

In DSR, routing overhead can potentially increase with increase in the number of hops between a source and destination station. Stale cache information can result in route errors and the chances for such errors are dependent of stations mobility patterns and the underlying channel quality which determines the local connectivity.

5.2.2.3 Temporally-Ordered Routing Algorithm (TORA)

TORA [15], uses proactive and reactive routing mechanisms. TORA assigns directions to links between stations (routers) in a multi-hop wireless network. The directions to the links are assigned by a station based on the metric values (heights) of neighbouring stations. The significance of associating heights with links is that a router can only forward packets to links that lead to lower (in terms of height-metric) routers. This avoids the formation of loops in the network.

Route discovery operation can be of two types: a reactively discovery route to a destination or a proactively maintaining a route by the destination. In the reactive mechanism, a source station initiates this process by sending a QRY message to the neighbouring stations. This message is propagated in the direction of destination station with every intermediate router determining links direction based on the neighbouring routers' height. Routers which have a valid route to the destination station replies with a UPD message to their neighbouring stations to help them determine the heights and link directions for that destination.

In proactive mode, a destination initiates routes creating by sending an OPT message to the neighbours. This message is forwarded by other stations and processed locally to calculate their own height with respect to the destination.

One of the requirements of TORA is the all stations should have synchronized timer.

5.2.3 Location based routing protocols

Location based routing protocols can enable the use of proactive routing protocols in a specific region of a network and rely on reactive routing protocols for further stations in the network. This brings the possibility to utilize the advantages of proactive routing and at the same time avoid the related overhead even in significantly large sized multi-hop networks.

Some of the significant contributions using location based routing approach are briefly discussed below:

5.2.3.1 Zone Routing Protocol (ZRP)

ZRP[16], divides the network into several zones. A routing zone for a specific station is constituted from the number of stations in its neighbourhood which are a particular radius (in terms of number of hops) from itself. The routing protocol in this zone is called the Intra zone routing protocol (IARP). The ZRP provides a modular design, therefore, a proactive routing protocol can be used for IARP [17, 35].

The routing protocol which is used for routing from outside a predefined zone is called Inter-zone routing protocol (IERP). A reactive routing protocol can be used for accomplishing IERP [18, 36].

5.2.3.2 Distance Routing Effect Algorithm for Mobility (DREAM)

DREAM [19] is a hybrid routing protocol as it uses a proactive method of dissemination routing information across the network and a reactive method of route selection. When a station wants to send a packet to a specific destination station in the network, (DREAM assumes that the current station would have the location information of the destination station), the current station sends the packet to its entire one-hop neighbours in the direction of the destination station. Each intermediate station also repeats the same procedure until the packet finally reaches its destination.

It is evident that DREAM relies heavily on the location information before selection of a route. It is important to work out a method of dissemination the location information throughout the network. However, dissemination of such information throughout the whole network can incur a higher routing related overhead. To control such overhead, DREAM uses the distance effect which means that the greater the distance between two stations, the slower they appear to moving with respect to each other. Therefore, location information can be updated in the routing tables as a function of the distance between stations. The distance effect is used in a way that highly mobile stations only distribute their location information to stations which are closer to it. Because, as the distance between these stations is short, the stations mobility would highly affect routing with the closer stations. DREAM controls the dissemination of such control packets by adaptively selecting the life-time of packets. When a station receives a packet, it calculates the distance that it has travelled through and decides whether to forward the packet by comparing the life-time value of the packet with the distance.

The frequency of updates in DREAM is determined by a station's mobility rate. For a highly mobile station, its location information needs to be updated in the network more frequently than static or in case of very low mobility stations.

5.2.3.3 Greedy Perimeter Stateless Routing (GPSR)

GPSR [20], uses location information for sending packets across a multi-hop ad hoc wireless network. It uses a *Greedy forwarding* and a *Perimeter forwarding* approaches. In greedy forwarding, a station marks a packet with its destination location. Every receiving station makes a local decision of further forwarding the packet. This decision is greedy in a sense that every station selects a neighbour for forwarding which is closest to the destination. The location information is disseminated by stations among its neighbouring stations using periodic beacons. The greedy forwarding method only relies on local information of one hop neighbours and their location.

In case of no stations in the neighbourhood which can use the greedy forwarding method, the GPSR uses perimeter forwarding method. In perimeter forwarding, a station selects stations along the perimeter of the void areas (where there are no stations for greedy forwarding) to route packets to the destination.

Geographic (location based) routing can potentially change the conventional routing mechanisms where every router keep a list of reachable destinations and the routes through which such destinations are accessible. In this case, the position information of the destination and the location information of the next hop neighbours can determine the route without any topological information.

However, the correct and on time dissemination of location information and the amount of control traffic which result from such updates presents a case where a trade-off can be made depending on the requirements of a specific MANET.

5.2.4 Routing using network clustering

Clustering is used to partition large size networks into small and well manageable groups. Clustering is achieved at the MAC and at the network layer [21, 22]. Clustering at the MAC layer has the benefits that the network capacity is increased by reducing the medium access delay. At the network layer, clustering can help in reducing routing overhead (which is the case for large size networks). The routing protocols which make use of network-layer level clusters are called hierarchal routing protocols. A challenge in designing hierarchal routing protocols is their ability to cope up with station mobility and management of the location information when stations move within different clusters. A number of cluster based (hierarchal) routing protocols have been proposed in the literature [23-26], in the following text we will briefly explain the Hierarchal State Routing (HSR) [23] protocol to provide an insight into the working mechanism of cluster based routing techniques.

5.2.4.1 Hierarchal State Routing (HSR)

HSR maintains a hierarchal topology of the network. In every cluster, a cluster head is elected by the other stations at the cluster. The cluster heads then re-elect another station to be the head of the cluster-heads and so on. The formation of clusters and election of cluster heads are dealt with the help of specialized algorithms, which generally use the policy of selecting a cluster head based on the computational resources, battery lifetime etc. In a cluster there are three entities, (1) the cluster head, which is responsible for managing all the transmissions within the cluster (2) the gateway station which is responsible for inter-cluster communication, and (3) and internal station. The aim of such hierarchal organization of the network is to reduce the network layer routing overhead in the form of reduced routing tables, processing and propagation of routing related control traffic.

There are two types of clustering supported by the HSR protocol: physical clustering, which depends on the physical location of a station within a cluster, and logical clustering, which depends on logical grouping of stations within clusters and is independent of the physical location of stations.

At the lowest level of the clusters all the stations monitor their connectivity with neighbouring stations and then communicate this information to the cluster-head. The cluster-head summarizes this information and sends to the neighbouring cluster-heads via the gateway station. In this way, new cluster heads are elected and the network connectivity information is propagated along the hierarchy.

The addressing scheme used in for hierarchal routing is also different and in HSR, a station's hierarchal address ID is formed as a sequence of the MAC addresses of all the nodes right from the top of the hierarchy to the node itself.

As discussed in this section, the MANET routing protocols in general are divided in to various categories, based on the mechanism in which the route-discovery process is accomplished, the type of addressing schemes used, the network overhead and the network hierarchy. There is not a single solution which fits in all network scenarios. A comprehensive comparison of various MANET routing protocols, targeting specific aspects of the routing protocols and their performance in various network scenarios are explored and reported in [27-34].

In this chapter, our focus is not to propose a yet another MANET routing protocol, but to focus on various issues which arise when local enhancements at the lower layers (specifically the medium access control sub-layer) are introduced, which directly or indirectly affect the performance of routing protocols. The aim of the research presented in this chapter is to derive a route selection metric, which truly reflects the state of the underlying layers.

5.3 MANET routing protocols- the constraints of MAC and capability of PHY of IEEE 802.11 conformant stations: *Analysis and Motivation*

In the IEEE 802.11 standard, broadcast frames are transmitted at the lowest data-rates [35]. The purpose of using the lowest transmission-rates is to enable larger coverage (given that the transmission rate is inversely proportional to the coverage area) and improve network connectivity. This feature of the MAC sub-layer has its implication on the performance of routing protocols for MANETs. Majority of MANET routing protocols use broadcasts during route discovery process. The route selection in most conventional routing such as Destination Sequence Distance Vector (DSDV), Ad-hoc On-demand Distance Vector (AODV) and Dynamic Source Routing (DSR), is usually based on the minimum hops between the source and destination. This is primarily a carryover from wired networks where the transmission rate of a link does not dynamically change and link rate is independent of the physical transmission range [36]. In a route with minimum hops between the source and destination, the distance between neighbouring nodes would be the highest and may cause creation of *communication gray zones* [37], where stations which are marked as '*reachable neighbours'* using the broadcast frames would not be accessible during normal, unicast (possibly using higher transmission rate) communication. Some factors which contribute in the formation of *communication gray zones* are: multi-rate transmission capability of the physical layer (PHY), no acknowledgment policy in the IEEE 802.11 MAC when broadcast frames are sent, the small size of broadcast frames which increases the transmission success probability of broadcast frames over unicast data (possibly larger) frames, and the link fluctuations which can disrupt border line stations. Rate adaptation schemes [38-48], exploit the multi-rate transmission capability of IEEE 802.11 PHY; and although it is not specified in the standard (until the IEEE 802.11n); it nevertheless plays an important role in improving the overall communication. To avoid the formation of *communication gray zones*, a routing protocol should be aware of the rate-adaptation at the PHY.

In addition to including transmission rate information in the routing protocols, it is absolutely important to include information about congestion (medium occupancy). In a Basic Service Set (BSS) area, having higher station density (and assuming that all are transmitting), a station may have to freeze its backoff counters a number of times before it finds an opportunity to transmit on the medium. Transmission errors and station density would account for higher medium occupancy by neighbouring stations and as a result every station would experience higher contention delay for accessing the medium. Likewise, stations using lower transmission rate can penalize neighbouring stations by causing higher medium contention. Therefore, a station may be able to transmit at a higher data rate; however, because of medium occupancy the medium access delay would be higher. In such situation, a route having higher throughput links would encounter higher end-to-end delay.

The end-to-end contention delay can potentially increase with increase in the number of stations along the route. Moreover, a routing protocol which selects a route by including links with higher transmission rates would generally imply that the route has a higher number of intermediate stations between a source and destination. Higher number of intermediate stations means relatively higher contention delay along the route. In addition, in a multi-rate multi-hop network, the end-to-end throughput is capped by a station using a relatively lower transmission rate along the route. If lower transmission rate links follow higher transmission rate links, frames would pileup in the transmission queues at a station transmitting at a lower rate and as a result there would be higher queuing delay [3]. The same is true when packets, on its way through the multi-hop route, land in a highly congested BSS. Avoiding, congested (having higher medium access contention delays) links and constructing a route with relatively less contention and queuing delays necessitates incorporation of MAC and PHY specific information in the routing protocol.

In summary, the MAC sub-layer is only responsible for handling single hop communication within a BSS area (where stations operate using a single/same coordination function). On the other hand a routing protocol is responsible for end-to-end delivery of packets. However, there are several MAC level factors such as the transmission rate, medium access contention and queuing delays, which are variable with time and have direct or indirect influence on the performance of a routing protocol. It is important to consider hop-by-hop (or a link) MAC level information while constructing an end-to-end routing strategy for a multi hop wireless network. In this chapter we formulate a routing metric, which combines variations in the transmission rates, contentions levels and queuing delay at every link. The novelty of the idea is twofold, in the first instance, the routing metric takes into consideration the variation of the queuing delay introduced prioritization of queues and frames within, this enhancements is added by the newer IEEE 802.11e MAC specification;

secondly, our routing protocol called MADV (see section-5.5), uses the proposed metric and uses per-AC routes per destination.

5.4 Related Research

In this section we present a rather long literature review of the work related to routing protocols which use link state information in constructing routing metrics. Despite the fact that we, to the best of our knowledge, precisely included the most relevant work, the length of this section is a reflection of the huge interest and contributions previously made in this area.

To complement the possibly unfair nature of the MAC sub-layer and its effects on multi-hop routing, there are several solutions [49, 50, 51], which propose multi-hop quality of service (QoS) guarantees during route establishment through resource reservation using the provisions in the IEEE 802.11e MAC. Such schemes provide end-to-end guaranteed service through an ad-hoc network. Reservation of resources along the route in a multi-hop network minimizes queuing (internal contention among traffic flows which belong to different traffic categories, and the flows are prioritized according to the IEEE 802.11e rules), and (to an extent) the external medium access contention. However, routing can still be affected by several other factors which cause fluctuation in link quality. In many cases such fluctuation may cause disruption of admitted flows (breakage of QoS routes), and route establishment in the case of QoS based routing is much more complex than best effort routing. Normally, in a QoS based routing, the resource reservation follows the process of route discovery (because resources cannot be reserved along a route which is not discovered, as there can be several candidate intermediate stations between the source and destination). Therefore, a routing protocol where route selection metric considers link quality parameters can be useful in QoS-guaranteed multi-hop routing protocols.

There are a number of '*best-effort'* MANET routing protocols which consider link state information (combination of transmission rate, contention delay, queuing, number of successful transmissions etc); *multi-rate aware sub-layer* (MAS) [52], modified DSR by introducing the capability to select a next hop node according to the transmission rate between two nodes. MAS keeps track of next hop neighbours state information such as the Signal to Noise Ratio (SNR) and the transmission rate between the two neighbours. This state information is maintained either by proactive (periodic HELLO message exchanges) or reactive (RTS-CTS-DATA-ACK exchanges) methods. [52] state that when the '*valid state*' bit is set, it implies that packet will be relayed to another node which is different from the node determined by the DSR. This node is called 'relay node' and the DSR has no information about it. Therefore, when the '*valid bit'* is set, and a packet it received at the *relay-node*, the *relay-node* does not forward the packet to its network layer. Although this approach enables dynamic adjustment of routes according to variation in the transmission rates. However, how relay nodes are discovered between DSR neighbours is not mentioned in [52], and the process of discovering relay nodes is essentially a *route-discovery process*, with the exception that the primary routing protocol is left unaware.

A similar mechanism is proposed in [53], where a station calculates standby time for forwarding RREQ proportionally to the medium time. The 'signal strength aware routing' (SSR) considers the transmission rate only at the time of route discovery, however, when MAC is rate-adaptive, the performance of SSR would be unpredictable. The authors also realized this shortcoming and they assumed 'no channel fluctuations' and 'no node mobility'.

Expected transmission count (ETX) metric is proposed in [54] which finds high-throughput paths on multi hop wireless networks. ETX minimizes the expected total number of packet transmissions (including the retransmissions) required to successfully deliver a packet to the ultimate destination. ETX, however, doesn't consider link congestions (medium access contention delays), and queuing delays. The forward and reverse delivery ratios are important in determining the ETX of a certain link, and the overall route. However, in this case the estimation of delivery ratios (in forward and possibly in reverse direction) use broadcast frames. In the standard IEEE 802.11 implementation, broadcast frames are sent at a lower transmission rate, which the stations may not use for transmission of unicast (data) frames. The delivery success probability of a frame sent at a lower rate (which use modulation scheme that is relatively robust to interference and other performance attenuation factors) is higher than frames which are sent at a higher transmission rate. Therefore, ETX estimated using broadcast frames will be over optimistic and thus inaccurate when (unicast) data frames are sent during normal communication. ETX relies on loss rate of links and does not consider the link transmission rate. As a consequence, a route with links operating at lower transmission rate which would (generally) have a lower ETX (because transmission failures are more likely to occur when a station operates at a higher transmission rate), would be selected. Similarly, ETX gives preference to shorter routes over longer routes as long as the ETX of the shorter route is less than that of the longer route. This again can lead to poor performance if the selected path (although shorter) has low transmission rate links.

To supplement the ETX metric, [55] proposed a bandwidth adjusted version of ETX, called Expected Transmission Time (ETT). ETT for a link is the time that a packet of certain *size* would take when transmitted at a certain *transmission rate* at ETX number of successful tries. The individual link weights are combined to form a path metric called *weighted cumulative ETT* (WCETT). The [55] uses ETX while calculating the ETT for a link and therefore, lacks the same features as ETX did (e.g. it is also uses broadcast frames for calculating the ETX, which proves to be over optimistic when MAC is enabled for rate-adaptation). ETT is designed for a network where all nodes are assumed to be stationary. Although, ETT is essentially a bandwidth adjusted ETX, the method of determining the bandwidth is not very accurate as pointed by the authors, and does not reflect the real conditions of a link. There is no clear evidence how the protocol will perform if a rate-adaptation scheme is used at the MAC sub layer which would dynamically select a different transmission rate according to the channel state information. Similarly, the ETT like the ETX doesn't include medium access contention delays at various links of a route. [56] presented a comparative analysis of various effects of MAC level enhancements on the performance of routing protocols. Rate-adaptation strategies such as ARF, RBAR and the rate adaptation as specified in the IEEE 802.11n standard specification were analyzed using ETT routing metric in the routing protocol. According to [56], ETT works well in providing high end-to-end throughput, however, it does not generate optimal throughput when it coexists with Aggregate-MPDU (A-MPDU) and EDCA with BACK.

The concept of ETX is also applied in [57] for calculating the Expected Data Rate (EDR) of a link. The EDR is then used to select a route in a reactive routing protocol.

The Medium Time Metric (MTM) [58] proposes a route selection metric using the transmission time, associated overheads and link reliability. A similar metric is used in [59] in a modified version of AODV.

MAC delay is used in [60] for calculating link costs and then the link costs are used to select a route in a modified version of AODV routing protocol. This approach has a number of shortcomings; for example it uses a fixed size packet, only consider the instantaneous transmission rate and the resulting MAC delay, it doesn't consider the medium access contention delay and the queuing delays at the intermediate nodes.

The authors in [61] studied the effect of rate adaptation mechanisms on the performance of routing and transport layer. According to their findings, ARF and shortest path routing protocols perform poorly, primarily because of the communication gray-zone problem. While on the other hand, rate-aware routing protocols can exclude stations which fall in the gray-zones during route selection and thus perform comparatively better.

The approach used by [36] considers almost every MAC level parameter which may have direct implication on the performance of routing. However, we feel that there are some aspects where further enchantments can be introduced e.g. while calculating the transmission time for packets, the number of retransmissions are assumed to be always 1 when a rate adaptation scheme is used. This assumption is not always true, because a rate adaptation scheme always tries to select the best transmission rate, and during this selection process a station may encounter a number of retransmission attempts [38-48]. Therefore, inclusion of the correct number of retransmission attempts while calculating the transmission time for packets may significantly affect the route selection. Likewise, various delay factors e.g. the IEEE 802.11 MAC specific inter-frame spaces and the backoff delays are omitted.

A congestion aware routing protocol using a metric which incorporates transmission rate, MAC overhead (medium access contention delay) and buffer delays in a variant of DSR is proposed in [3]. During the route discovery a station would only forward the RREQ messages if the 'effective link data-rate category' ELDC of the link is greater or equal to that specified in the RREQ. This essentially means that the routing protocol would ensure that the selected route is capable of ELDC throughput (which in this case is 6). All the MAC level parameters which influences the performance of a routing protocol are variable, however, [3] uses MAC specific information only in the beginning, i.e. at the route discovery time.

An upper bound for the achievable throughput using the 802.11 MAC, considering multi-rate capability of wireless stations is modelled in [62]. Using that model [62] formulated a contention aware transmission time routing metric for OLSR. The mechanism for calculating a route metric requires the availability of a number of information about the neighbouring nodes, e.g. there has to be information about size of packets used for transmission by the neighbouring nodes along with the transmission rates used during the transmission. The packet size information from a neighbouring station may not be easily available (although it can be found indirectly, from the NAV and using the transmission rate information used for transmitting a particular frame), and therefore, it may complicate the estimation.

5.5 Medium Aware Distance Vector (MADV) routing protocol

5.5.1 Medium Awareness

The hierarchal location of routing protocols in the layered protocol architecture depends on the addressing scheme which is used by the routing protocol while determining the routes between a source and destination station in a network. The main objective of using layered architecture for designing the communication system and positioning the corresponding protocols at a certain layer is to provide various levels of abstractions among the layers. It essentially means that a protocol operating at layer 'X' should not be concerned with the functionality of the underlying layers/or its associated protocols. Such levels of abstraction make it very easy to completely replace/upgrade a certain set of protocols at a specific layer without any changes to other layers.

However, in case of wireless networks, the behaviour of the Physical layer in terms of transmission rate, frame error rate and reliability is not consistent; it is dependent on a number of factors each of which play their own proportionate role. Likewise, the design of MAC protocol, which is essentially based on the Carrier Sensing Multiple Access with Collision Avoidance (CSMA/CA), internally induces its own constraints. Such constraints depend on the number of participant stations in the vicinity, channel quality, and the type of traffic streams. Variations in the performance of underlying layers' performance makes it highly complex to define the 'State of the medium'; and due to the involvement of so many factors, which affect the communication, it becomes mandatory to associate a list of parameters when it comes to defining the state of medium. Likewise, due to the interdependencies, and implication of variation of these parameters on the performance of protocols which are resident at higher layers, it seems logical to increase cooperation between the layers. Such cooperation, which is widely termed as cross-layer design, may result in blurring the boundaries of layer-to-layer abstraction. Cross-layer design is deemed to be necessary in the case of emerging wireless networks and have been reported to enhance the overall efficiency [4-8].

In connection with the preceding discussion, it is logical to define the 'state of the medium' in terms of all the parameters which affect the performance and efficiency of communication. A mechanism of monitoring such parameters would create the medium-awareness function.

5.5.2 Medium monitoring parameters

In the perspective of MADV-routing protocol, medium awareness includes parameters from the MAC and PHY layers. At the MAC layer, parameters of interest are: (1) the disproportionate queuing delays of various traffic classes/Access categories, and (2) the contention delays before the transmission of every frame. The PHY layer parameter which is important for consideration in MADV includes the average transmission rate (which is essentially converted to the average transmission time) over a certain averaging time window.

It is important to notice that frame error rate (due to interference- corrupted frames, and due to MAC failure resulting in simultaneous transmission- collided frames) also plays an important role as a PHY-level parameter. However, in the case of MADV, the routing protocol assumes to have a rate-adaptation module at the MAC layer. The rate-adaptation module's prime functionality is to monitor the CSI (which includes the both variants of erroneous frames, thus including the frame error rate) and adaptively select an appropriate transmission rate. In essence, using the output of a rate-adaptation module, which in this case is the average transmission rate, would reflect the frame-error rate during that average time window through the variation in transmission rates. The medium monitoring parameters as used in the MADV are explained as follows:

5.5.2.1 Queuing delay (Tq)

Queuing delay represents the time spent by a packet in the MAC sub-layer queues (where it is called MAC Service Data Unit (MSDU), after passing through the MAC-service access point), till it is marked 'ready' for transmission. The arrival and service rate of the queue affects the queuing delay.

Packet arriving in the queue are from two different sources, (1) packets received from other stations which are further relayed and (2) are packets generated by host applications. Service rate is tied up with the channel conditions, transmission rate, contention levels within a BSS and the internal contention within the queues at every station.

The IEEE 802.11e[9] defines four access categories (ACs) i.e.

(1) AC BK with the lowest priority, where the type of service is 'background'.

(2) AC BE having higher priority than AC BK, and type of service it is associate is 'best effort'.

(3) AC VI having higher priority than AC BE, and type of service is 'Video'.

(4) AC VO having the highest priority, and the type of service it is associated is 'voice'.

Packets are tagged by applications to associate them with a particular class of delivery service; the tagged packets, at the MAC layer fall in either one of the four listed categories of AC queues. Priority among frames in the four queues is accomplished by using different values for the CW_{min}, CW_{max}, AISFN, and different TXOP Limits (see Chapter-2 for details). Unlike the MAC layer operation, when two stations access the medium simultaneously and result in collision, no one gets access to the medium; in case of the internal queues, when there is an internal collision, the priority is always given to a higher AC. Prioritizing frames on the basis of ACs creates unfair medium access distribution among frames. This essentially means that in the presence of higher priority frames in the queues, the lower priority frames would have to wait for a longer time at every station before accessing the medium.

5.5.2.1.1 Significance of queuing delay

It is important to analyze the significance of the queuing delays at the MAC level. In order to appropriately estimate the queuing delay, it seems logical to consider the queuing delay experienced by frames in the lowest priority queue (AC BK). For such analysis, a series of simulation scenarios were configured in OPNET [63].

In the scenarios, IEEE 802.11e compliant transmitter-receiver pair of stations were configured in a client-server relationship. The server provided services for file sharing using FTP, Database access, Video streaming and Voice. At the MAC level, the EDCA parameters configuration for all four ACs is given in Table 5-1.

[9] A QoS IBSS supports operation under the HCF using TXOPs gained through the EDCA mechanism. The parameters that control differentiation of traffic classes using EDCA are fixed.

Table 5-1: EDCA parameters configuration for the four Access categories

Access Categories	CW_{min}	CW_{max}	AISFN	TXOP Limits
AC_BK	aCW_{min}	aCW_{max}	7	One MSDU
AC_BE	aCW_{min}	aCW_{max}	3	One MSDU
AC_VI	$(aCW_{min} - 1)/2 - 1$	aCW_{min}	2	3008 µs
AC_VO	$(aCW_{min} - 1)/4 - 1$	$(aCW_{min} - 1)/2 - 1$	2	1504 µs

Four applications were configured. In each of the scenarios the delivery service was changed for each of the application. In scenario 'AC_BK', two of the four applications were enabled to generate traffic. The type of delivery service in this scenario for both applications was set to 'background' which is translated at the MAC level to the AC_BK. Therefore, the outbound frames in scenario 'AC_BK' would be queued in the in AC_BK queue.

In scenario 'AC_BK + AC_BE' one of the application used delivery service of 'Background' and another used 'Best Effort'. Therefore, it is likely to see the AC_BE traffic prioritized over the AC_BK. This is shown in Figure 5-1; where it shows the comparison of MAC level delay (Queuing + Medium access) for frames in the AC_BK Queue, in two scenarios, Scenario: AC_BK shows the MAC level delay when all the outbound frames are tagged with AC_BK delivery service, while, scenario: AC_BK + AC_BE shows the MAC level delay when the outbound frames are of belong to two ACs i.e. AC BK and AC BE. It is evident that delay experienced by AC_BK traffic is higher in the 'AC_BK+AC_BE' scenario owing to higher priority traffic at the MAC layer.

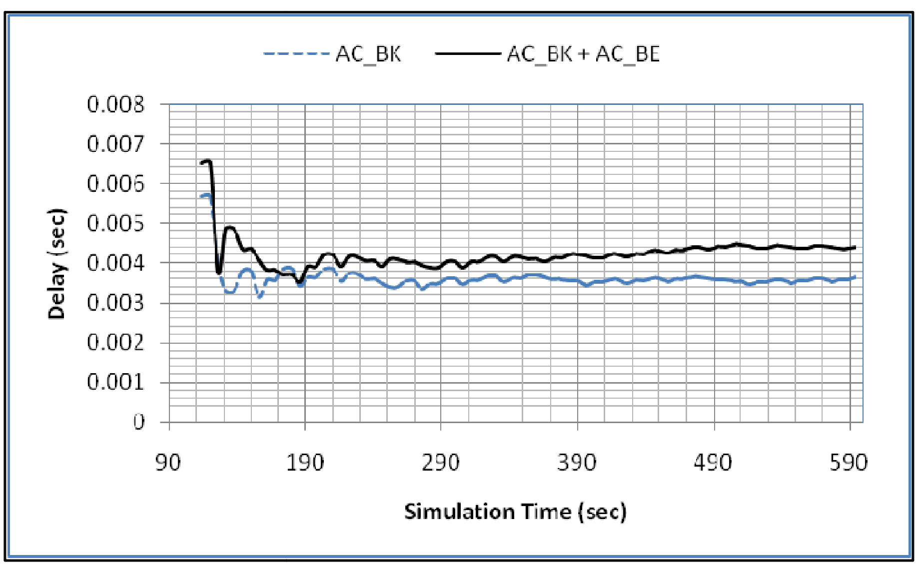

Figure 5-1: Comparison of Queuing and medium access delay for frames in the AC BK Queue in two scenarios.

Figure 5-2 shows the comparison of MAC level delay (Queuing + Medium access) for frames in the AC BK Queue, in four scenarios. Scenario: 'AC_BK' shows the MAC level delay when all the outbound frames are tagged with AC BK delivery service, while, scenario: 'AC_BK + AC_BE' shows the MAC level delay when the outbound frames belong to two ACs i.e. AC BK and AC BE. Scenario: 'AC_BK+AC_BE+AC_VI' and Scenario 'AC_BK+AC_BE+AC_VI+AC_VO' shows the MAC level delay when the outbound frames belong to three ACs and four ACs.

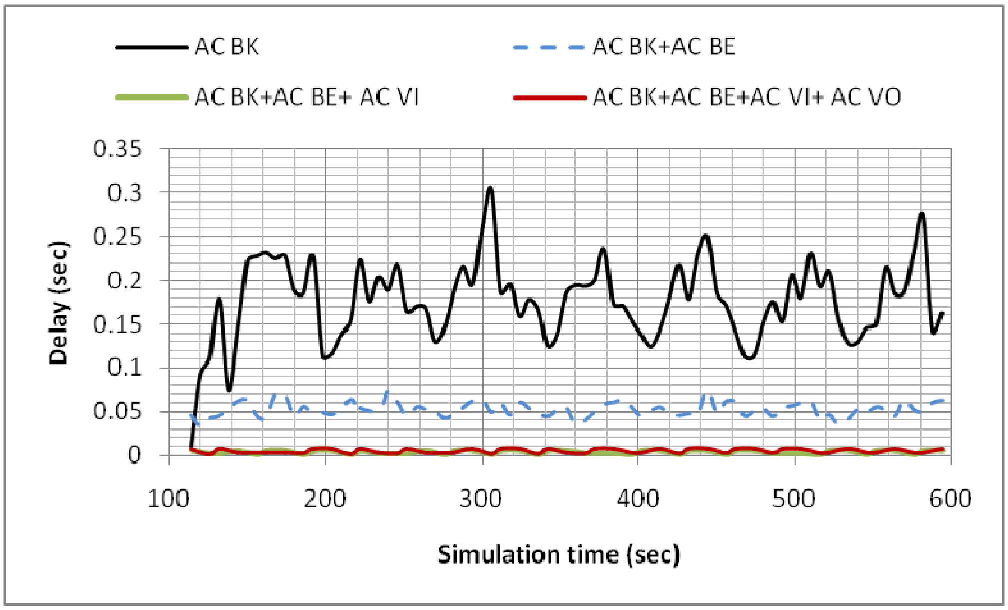

Figure 5-2: Comparison of Queuing and medium access delay for frames in the AC BK Queue in four scenarios.

When frames from higher priority ACs are present at the MAC layer queues, the frames from lower priority ACs suffer higher degree of internal collisions. This is shown in Figure 5-3, where the statistics collection mode which is used here is 'Bucket', with a bucket size of 100 frames and the statistics recorded is sum of the bucket size. The values in this figure essentially represent the percentage of internal collisions for frames belonging to AC_BK queue in the scenarios. In the AC_BK scenario, there are no internal collisions and that is why it is not shown in Figure 5-3.

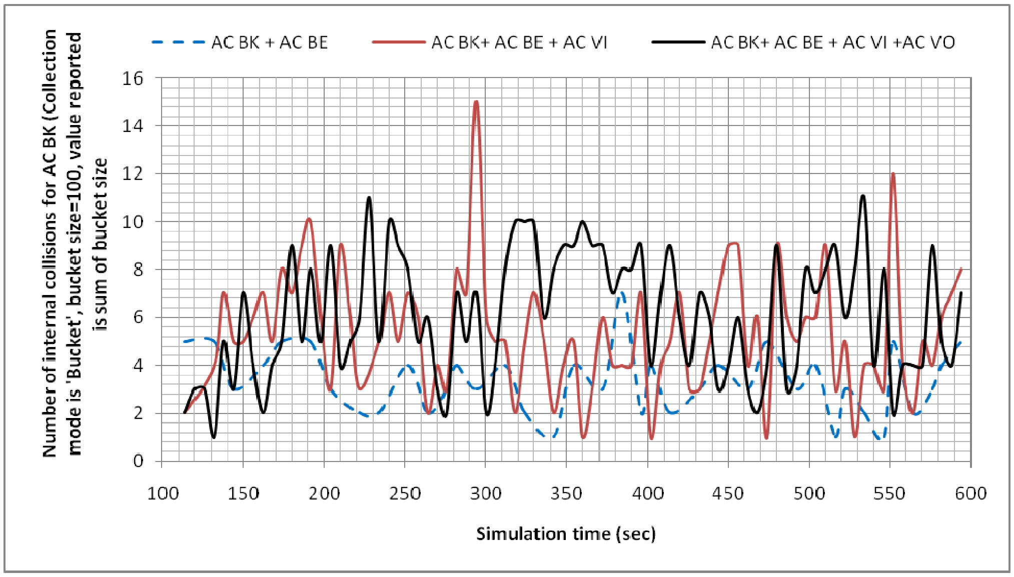

Figure 5-3: Comparison of internal collisions for frames belonging to AC_BK queue, in the four simulation scenarios.

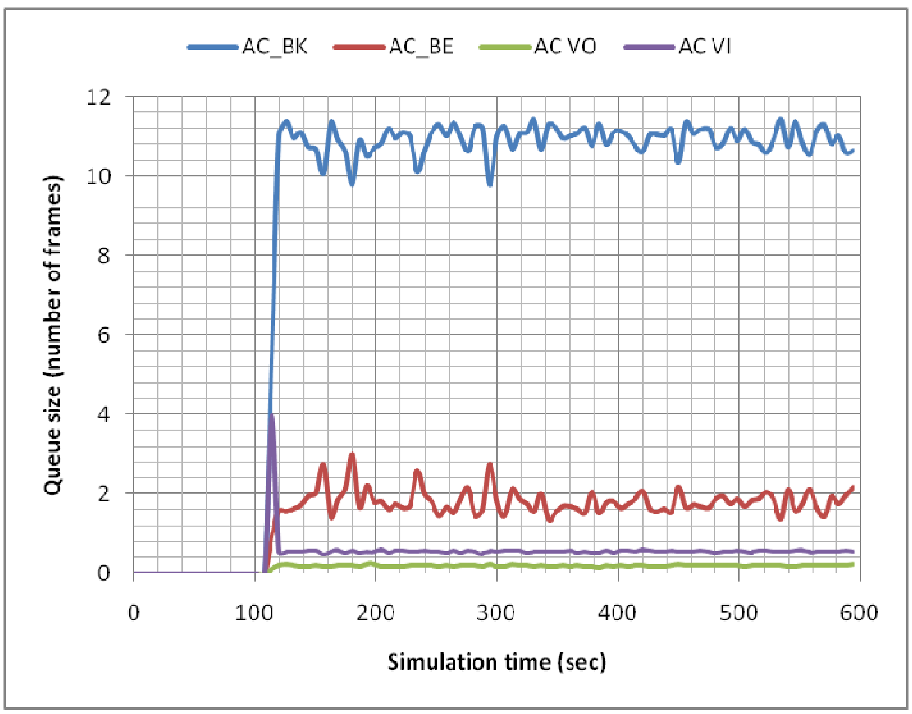

Figure 5-4: Queue sizes of four different ACs queues (AC BK, AC BE, AC VI and AC VO) using a similar traffic arrival rate in each queue.

Prioritization at the MAC layer introduces uneven queuing delays. To demonstrate the impact of this phenomenon, we collected the queue-size statistics for each of four ACs in the scenario 'AC_BK+AC_BE+AC_VO_ACVI'. Due to variation in the service rate (introduced by the prioritization) the queue size for AC_BK grows higher. This is shown in Figure 5-4.

As a result of the internal contention within the queues (which causes uneven delays), every AC would have its own view of the network and have different routes even to the same destination station because of different values of the routing metric. This possibility necessitates that at a station's routing table; there should be different routing entries (one) for every AC to a destination. The structure of MADV routing table is discussed in section 5.5.6, and an instance of MADV routing table is given in section 5.5.7, in Table 5-3.

5.5.2.2 Contention delay (Tc)

Contention delay is the time that a station spends while competing for acquiring access to the medium. The contention period starts when an MSDU is marked 'ready' for transmission. Under the IEEE 802.11 rules a station waits when medium is busy. When the medium becomes idle, a station further waits for predefined distributed inter-frame space (DIFS) duration and then waits for a random number of timeslots depending on the value of back-off. All stations freeze their back-off counters when there is an activity on the medium.

The value of back-off counter depends on the outcome of previous transmissions; retransmissions result in higher values of back-off counters (by incrementing the Contention Window size). In a highly congested BSS, stations would freeze their back-off counters a number of times before they finally find their turn for transmission. To demonstrate the significance of backoff and the overall medium access delay we simulated various scenarios with varying station density and configured the stations to communicate whenever they acquire access to the medium. Figure 5-5 and Figure 5-6 show the backoff delay and contention delay variation respectively, when the number of stations are increased.

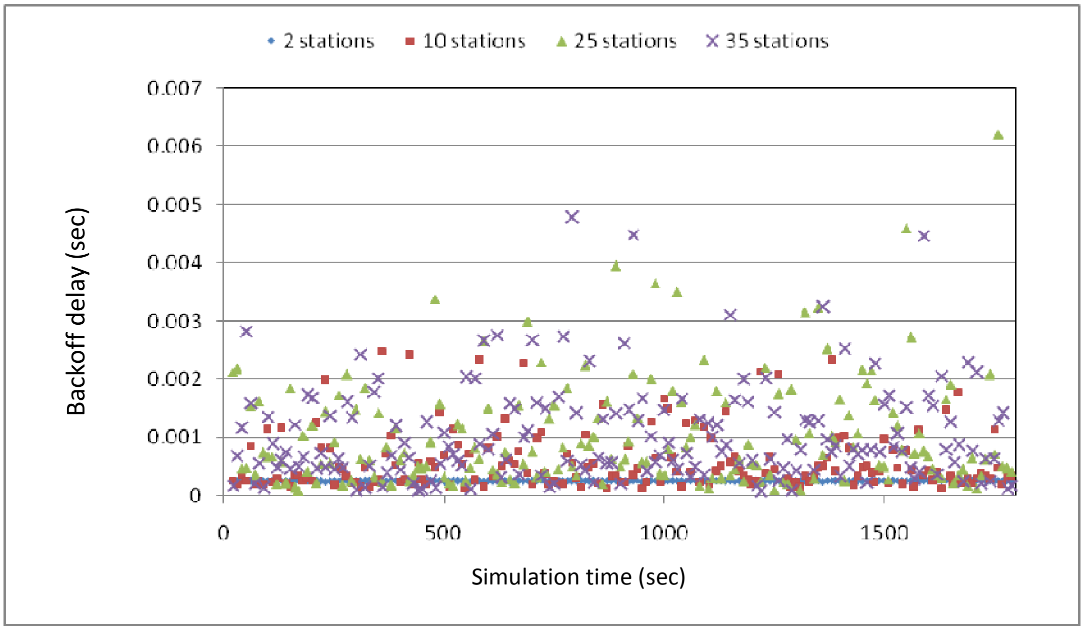

Figure 5-5: Backoff delay increases with increase in the number of stations in a BSS. Scenarios with 2 stations are compared with three other scenarios with 10, 25 and 35 stations. The reason for higher backoff delay is that the probability of simultaneous medium access by more than one stations increases with increase in the number of stations.

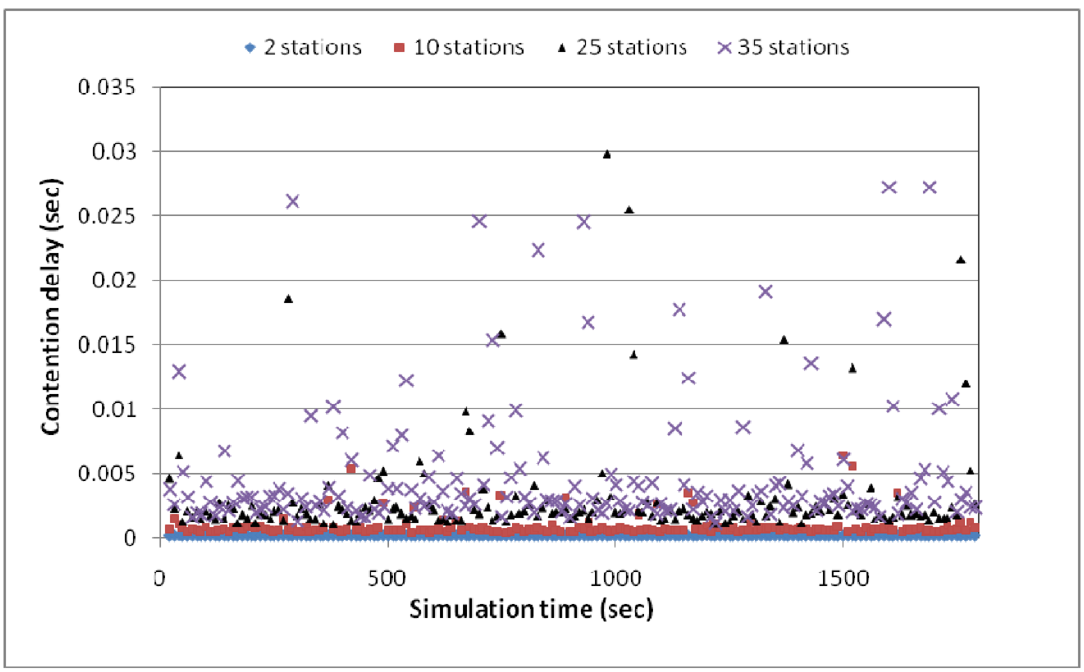

Figure 5-6: Medium contention delay increases with increase in the number of stations in a BSS; it also depends on the traffic patterns and the transmission rates used by the stations. Scenarios with two stations are compared with three other scenarios with 10, 25 and 35 stations. The stations used CBR traffic and used a fixed transmission rate.

5.5.2.3 Transmission delay (T_t)

The parameter 'T_t' represents transmission and propagation time for delivery of MSDUs between MAC level peers (on two stations), including the time spent in the retransmissions.

MAC level retransmissions significantly affect the decisions of statistics-based rate adaptation schemes [38-48]. For example, Auto Rate Fallback (ARF) [38] and Adaptive-ARF [40] decreases the transmission rate after two frame failures, and increases it after when the number of successful transmission crosses a threshold; in case of ARF the threshold is 10 while AARF uses an adaptive value. Therefore, after frame failures, it is likely that retransmission attempts would be made using a lower transmission rate, which would cause relatively higher transmission delay. Our implementation records the per frame T_c and T_t as follows:

$$T_c + \mathrm{T_t} = \sum_{\mathrm{n=1}}^{\mathrm{Retrans}} (\mathrm{T_c} + \mathrm{T_t})_{\mathrm{n}} \quad (5.1)$$

In other words, every value of T_c and T_t represents the total contention and transmission delay experienced till a frame is successfully transmitted.

5.5.3 Routing metric

The routing metric of MADV at a station represents the sum of the average queuing, medium access contention and transmission delays experienced by a frame when sent to a destination. For a destination station '*x*', a station running MADV would have the following route metric in its routing table:

$$\mathrm{MADV}_{\mathrm{RM}} = \sum_{\mathrm{i=1}}^{\mathrm{TNH}} (\mathrm{Tq_i} + \mathrm{Tc_i} + \mathrm{Tt_i}) \quad (5.2)$$

Where, Tq, Tc and Tt are queuing, contention and transmission time at each link of a route to destination 'x' respectively. 'TNH' is the total number of hops between a source and destination stations.

5.5.4 Realization of MADV routing metric as a routing protocol

MADV is an alternative approach of using the underlying MAC-layer parameters in construction of the routing metric. In order to analyze the performance of MADV's routing metric, it is important to devise a complete routing solution for an ad hoc wireless network which will essentially include route-discovery, route-maintenance, routing-error correction and propagation, and route-update mechanisms. Fortunately, due the tremendous efforts during the last two decades, there are several routing protocols, for mobile ad-hoc wireless networks, belonging to various classes of routing protocols. A brief summary of various classes of routing protocols is given in section-5.2.

Our objective with the use of MADV is not to implement yet another routing protocol for routing in mobile ad-hoc multi-hop networks, but to integrate the MAC layer enhancements into the routing protocols in the form of the MADV metric for performance enhancements. As discussed in the 5.3, standalone MAC layer enhancements when not taken into consideration can adversely affect the performance of routing protocols. MAC layer is the closest level to monitor the channel state information of a time-varying quality of channels. We believe that by integrating the MAC layer parameters, which are significantly important for determining the quality of communication in terms of various QoS parameters (e.g. delay, throughput, frame error rate etc), we can significantly improve the performance of existing routing protocols.

The MADV-metric can be incorporated into various routing protocols and its applicability is determined by the possibility of provision of MAC dependent Q.C.P parameters which are used to determine the hop by hop MADV-metric values.

For analysis of the potential of MADV-metric, we incorporated the MADV-metric in AODV routing protocol.

5.5.5 Decoupling network performance from hop-count

MADV monitors the underlying network's performance by using the Queuing, Contention and Propagation delay (Q.C.P) parameters. For every pair of stations (neighbours) 'x' and 'x+1':

1. The contention delay $'Tc'$, for all four access categories would be same. A station 'x' would experience the same $'Tc'$ while communicating with all its neighbours.

2. Within each queue of the four access categories queues, a station 'x' would experience the same queuing delay while communicating with its neighbours.

3. For every neighbouring station, a station 'x' would (possibly) experience a different propagation delay, due to the variable transmission rate which may (possibly) be in use for transmission to the neighbouring station.

Queuing and contention delay is dependent on the selection of appropriate *stations* along the route which offer minimum queuing and contention delays. On the other hand, minimizing the propagation delay requires the selection of *links* which offer minimum propagation delay. In both cases, the number of hops is completely irrelevant for improving performance of multi-hop communication. Therefore, it is important to decouple the hop-count metric which is conventionally associated to best path, by integrating the MAC layer parameters into the routing protocol.

5.5.6 MADV route entries

We modified the structure of AODV routing table by including the Q.C.P parameters as route entry parameters. In OPNET, AODV routing table is collection of route entries, where each route entry is declared as a C-structure called 'AodvT_Route_Entry'. In the AODV routing table, the route entries are created based on the destination address and consists of several fields, e.g. *destination sequence number, a (binary, Boolean) flag showing the validity of destination sequence number, next hop address, next hop port information, hop count, pointer to a list of precursors (which are served by the current station for the specific destination), route expiry time and* other fields for the internal processing by the simulator.

For the MADV, the route entry structure[10] has four next-hop address fields for the same destination address, each of the four correspond to a one of the four AC-routes. The requirement for creating separate next-hop fields, and the formation of AC-routes is explained in the following section (please see Section 5.5.7).

Depending on the metric values, for a unique destination address, a station running MADV may possibly have a different view of the network for each of the four access categories. This implies that if route-A is selected to be the best route for forwarding packets which belong to AC_VO, the same route may not be ideal for packet belonging to AC_BE, even for routing between the same source and destination stations.

The MADV routing table has such a provision for maintaining per AC route information. An instance of the structure of MADV routing table is as given in Table 5-2. The entries shown in this table are arbitrary values for the purpose of explanation.

Table 5-2: An instance of MADV routing table.

Destination Address	AC_BK		AC_BE		AC_VI		AC_VO	
	Next hop for QCP_AC_BK	MADV_QCP for AC_BK	Next hop for QCP_AC_BE	MADV_QCP for AC_BE	Next hop for QCP_AC_VI	MADV_QCP for AC_VI	Next hop for QCP_AC_VO	MADV_QCP for AC_VO
192.168.100.3	192.168.100.15	0.06	192.168.100.15	0.055	192.168.100.13	0.04	192.168.100.12	0.035

5.5.7 MADV operation: *an overview.*

To understand the MADV protocol, consider the topology in Figure 5-7. This topology will be used for reference in section 5.5.8.1, and section 5.5.8.2. Station 1 (S1) communicates with S9 through intermediate stations in an ad-hoc network. The values at each station in Figure 5-7 represent the link metric values (where each given value is a sum of Q.C.P delay) for AC_VO, AC_VI, AC_BE and AC_BK respectively as shown in Figure 5-8. For

[10] typedef struct

```
{
IPT_Dest_Prefix         dest_prefix;
int                     dest_seq_num;
Boolean                 valid_dest_sequence_number_flag;
InetT_Address           next_hop_QCP_AC_BK;
InetT_Address           next_hop_QCP_AC_BE;
InetT_Address           next_hop_QCP_AC_VI;
InetT_Address           next_hop_QCP_AC_VO;
double                  MADV_QCP_AC_BK;
double                  MADV_QCP_AC_BE;
double                  MADV_QCP_AC_VI;
double                  MADV_QCP_AC_VO;
List*                   precursor_lptr;
.
.
.
} MADV_Route_Entry;
```

example, S1 has a link metric of 0.5 for AC_VO when it communicates with S3. The link metric values as shown in Figure 5-7 are arbitrary numbers (for simplicity) and do not represent actual values. Table 5-3 shows the forwarding table of S1, showing entries for S9, S7, S4 and S8.

The link metric values are higher for AC_BE and AC_BK at S3, S5 and S7 which may reflect the possibility that these stations either generate or forward packets most of which belong to either AC_VO or AC_VI. As a consequence packets which fall into other ACs would experience higher queuing delays at these stations. According to the forwarding table of S1 and the metric values as depicted at various stations in the topology shown in Figure 5-7, the routes used by various ACs from S1 to S9 are:

AC_VO:	S1→S3→S6→S8→S9
AC_VI:	S1→S3→S5→S8→S9
AC_BE:	S1→S2→S4→S7→S9
AC_BK:	S1→S2→S4→S7→S9

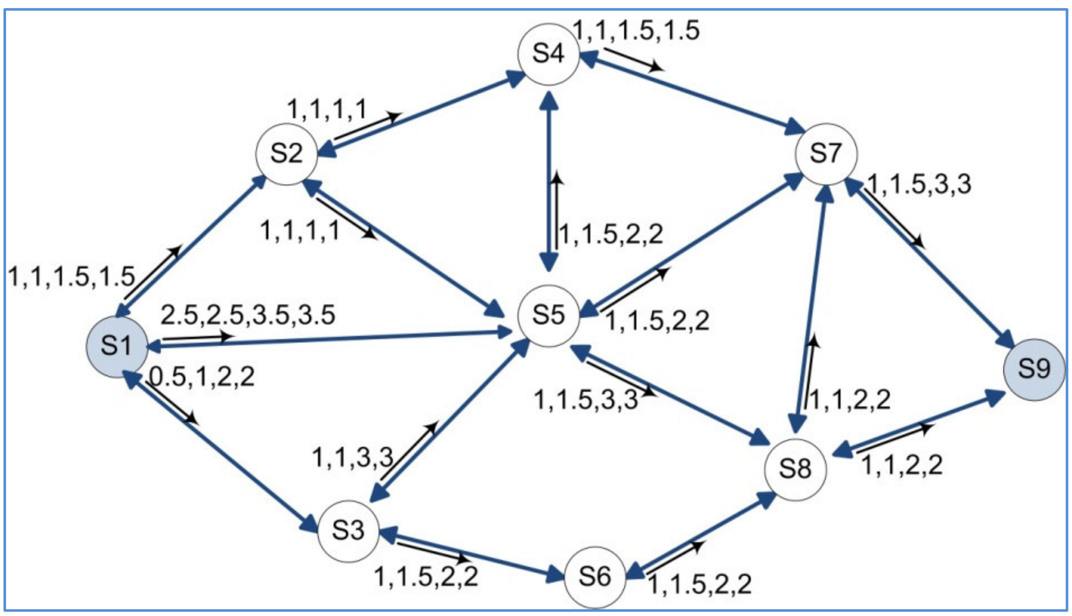

Figure 5-7: An Ad-Hoc network showing link metrics for individual ACs at every station.

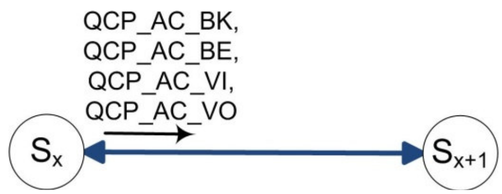

Figure 5-8: MADV route metric for a link between two stations.

Table 5-3: MADV routing table for topology shown in Figure 5-7.

Destination Address	AC_BK		AC_BE		AC_VI		AC_VO	
	Next hop for QCP_AC_BK	MADV_QCP for AC_BK	Next hop for QCP_AC_BE	MADV_QCP for AC_BE	Next hop for QCP_AC_VI	MADV_QCP for AC_VI	Next hop for QCP_AC_VO	MADV_QCP for AC_VO
s9	s2	7	s2	7	s3	4.5	s3	3.5
s7	s2	4	s2	4	s3	3.5	s3	2.5
s4	s2	2.5	s2	2.5	s2	2	s2	2
s8	s2	5.5	s2	5.5	s3	3.5	s3	2.5

5.5.8 MADV route discovery and maintenance

The MADV uses the same approach as AODV for route discovery and maintenance, with slight modification for integrating the new route metric for route selection. Route discovery is performed with the help of route request (RREQ) and route reply (RREP) initiated by the source and destination/intermediate stations respectively.

For inclusion of the MADV route metric, a station needs to know about the current contention delay experienced at its MAC layer. The station also needs the average queuing delay information for each of the four ACs. These two parameters i.e. the contention and queuing delay is specific to the current station and does not depend on the next hop (or the specific neighbouring station). On the other hand, for inclusion of the link propagation delay, a station needs to have the record of the transmission rates which are in use for each of its neighbouring stations. In case of no rate-adaptation at the MAC layer, a station would use fixed transmission rate for communication with every station in its neighbourhood. But using a fixed transmission rate leads to inefficient utilization of wireless spectrum and inefficiency in the overall communication (See chapter-2 for detail of rate adaptation schemes). Using the fixed transmission rate would also reduce potential of MADV; where, it would only be able take into account the queuing and contention delays.

Figure 5-9 shows the Q.C.P parameters, where the propagation delay from a station 'x' to its neighbours 'x+1', 'x+2' and 'x+3' is different, indicated by $P_{x \rightarrow x+1}$, $P_{x \rightarrow x+2}$ and $P_{x \rightarrow x+3}$respectively. While the contention delay and queuing delay for station 'x' is dependent on the MAC layer of the station itself and thus indicated by C_x and Q_x respectively.

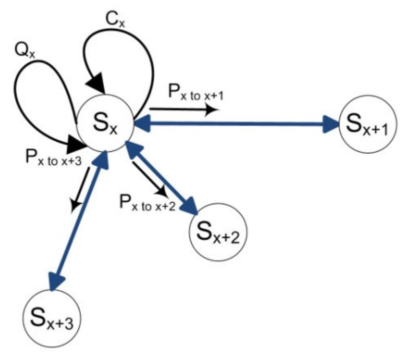

Figure 5-9: MADV local MAC parameters.

5.5.8.1 RREQ propagation and processing

The MADV modifies the original format of the AODV RREQ frames to include:

(1) Queuing delay from source to the destination/intermediate node handling the RREQ for all four Access Categories (ACs),
(2) Contention delay and
(3) The Propagation delay fields.

From here on queuing, contention and propagation delay fields would be refereed as Q.C.P-delay fields. The objective of using the Q.C.P-delay fields in the RREQ message field is to allow the destination station/or a station processing the RREQ message, to have the route-metric values of the reverse route back to the source (The inclusion of Q.C.P-delay field is explained in detail later in this section).

A station initiating a RREQ message assigns *NULL* values to all the Q.C.P-delay fields. If the station receiving the RREQ is not the intended destination but has a valid route to the destination; it simply fetches the entry from its routing table using the destination prefix. The Q.C.P-delay values in the routing table of the current node shows the corresponding cumulative delay values for Q.C.P-delay from the current station to the destination station (for more on how a station maintains its routing entries, please see RREP processing). Therefore, a RREP message is sent back to the station sending which sent the RREQ and the precursor list is updated. The Q.C.P-delay fields in the RREP message indicate the Q.C.P-delay from the current station to the destination station.

If the station receiving the RREQ is the intended destination; after receiving the RREQ, it updates its routing table if either of the Q.C.P-delay fields shows lower values than the corresponding values it has in its routing table for the source station. If there is no previous entry in its routing table, it simply inserts a new entry. The destination station sends back a RREP message (see RREP propagation and processing).

If the station receiving the RREQ is not the intended destination and its routing table has no entry for the destination prefix, the station has to rebroadcasts the RREQ message, adhering to the rules of AODV i.e. checking if the hop-count has not crossed the Network-diameter value, etc. Before, the rebroadcast of an RREQ message, a station adds its Q.C.P-delay values to the corresponding fields in the received RREQ message. It is important to notice that the propagation delay value at this stage is calculated based on the transmission rate between the current station and the station which sent the RREQ message; see

Figure 5-10, where the station s4 adds its own queuing delay (Q_4), contention delay (C_4) and the propagation delay between itself and the station which sent the RREQ message which in this case is 's2', therefore the propagation delay included by 's4' is P_{4-2}.

The current station after receiving an RREQ message, first checks if it has recent (in the past 10 seconds time-window) communication with the station sending the RREQ message. It is important to notice that the propagation delay calculated by 's4' is not based on the transmission rate at which 's2' sent the latest RREQ message; because as mentioned in section-5.3, RREQ is a broadcast frame which is sent by the MAC layer of IEEE 802.11 at a lower rate than the transmission rate which is used for actual transmission of data frames. A rate-adaptation module maintains a local cache of neighbouring stations (MAC addresses) and the transmission rates used for communication with each of the neighbouring stations. It is also possible for the station to maintain a similar field in the local cache to indicate the transmission rates used while communication with the station which transmitted the RREQ-message. Therefore, if there is a recent communication with the RREQ-sending station, a simple table look up (fetching the transmission rate using the source MAC address) provides the transmission rate information of the station which transmitted the RREQ message. A station receiving a RREQ message fills the propagation delay field by using the transmission rate information from its local tables. However, in case of no recent communication between the stations sending and processing the RREQ; the receiving station uses a default transmission rate of 11 Mbps and using this rate it estimates the propagation delay. Considering the propagation delay from the current station to the RREQ-sending station provides the accurate measure of the propagation delay which would be experienced on this hop of the route.

An RREQ message which is propagated in this manner provides information about the route from the destination to the source station in terms of the selected metric parameters. A destination station/or any intermediate station which responds to an RREQ message may receive more than one RREQ messages from neighbouring stations, and can select a route back to the source based on the values of the route-metric parameters. The following example, in

Figure 5-10 will elaborate on the RREQ propagation and processing mechanism used in MADV. This figure shows the processing of RREQ message along the route s1→s2→ s4→s7→s9. The grey boxes shows the MADV metric processing at the source and intermediate stations. The white boxes at the destination shows the MADV route metric at station s9 for destination s1.

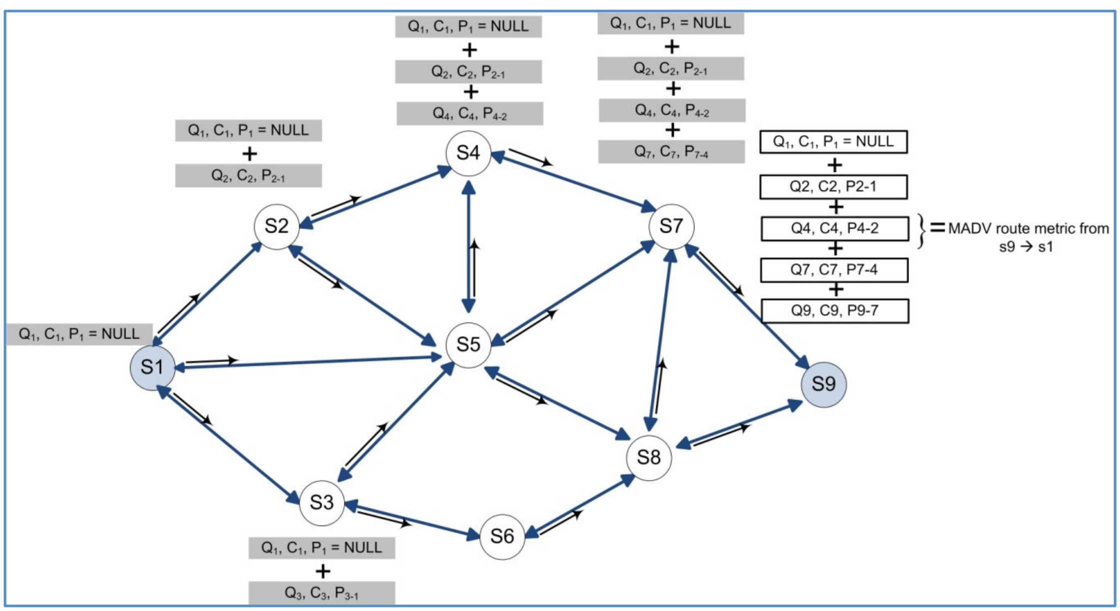

Figure 5-10: MADV RREQ propagation and processing.

5.5.8.2 RREP processing

A RREP message is generated in response to a RREQ message. A RREP indicates the queuing, contention and propagation delay from the destination to the source.

A destination station after receiving a RREQ message, increments its sequence number if the destination sequence number in the received RREQ message is higher than its current sequence number to match the value. Otherwise, it does not change its sequence number. The destination station places the sequence number in the corresponding filed in the RREP message and assigns NULL to the QCP parameters.

If the current station generating the RREP message is an intermediate station, it simply places the destination sequence number that it has maintained for the destination in to the corresponding filed in the RREP message. The intermediate station also updates its precursor's lists to include the source of RREQ message. The intermediate station, fetches the QCP parameters' values that it has for the destination and adds its corresponding QCP values for inclusion in the RREP message. The rest of processing remains the same as in the case of AODV i.e. the lifetime field in the RREP is calculated by subtracting the current time from the route expiration time value that the intermediate station has for the current route.

A source station may receive more than one RREP messages for the same destination station, each of the RREP messages may possibly give a different value of the route metric ($MADV_{RM}$). The source station selects a route based on the comparison of the received values of $MADV_{RM}$ in the corresponding RREP messages. It is important to notice that such comparison is made individually for each AC. For instance, consider the topology in Figure 5-11; assume that the source station receives two RREP message indicating two different route metrics (i.e. $MADV_{RM1}$ and $MADV_{RM2}$), we simply name the route: s1→s2→24→s7→s9 as route-1 and route: s1→s3→26→s8→s9 as route-2. Assume that the route-2 offers higher queuing delay for traffic category belonging to AC_BK, and less for traffic belonging to AC_VO and AC_VI such that:

$$(Q_{AC\_BK}+C+P)_{route\text{-}1} < (Q_{AC\_BK}+C+P)_{route\text{-}2}$$

$$(Q_{AC\_VO}+C+P)_{route\text{-}2} < (Q_{AC\_BK}+C+P)_{route\text{-}1}$$

$$(Q_{AC\_VO}+C+P)_{route\text{-}2} < (Q_{AC\_BK}+C+P)_{route\text{-}1}$$

In this case, the station ‘s1’ would use route-1 for packets tagged with delivery service ‘Background’ and update its routing table entries at time instance t1 (when it receives RREP for route-2), to use route-2 only for packets tagged with ‘AC_VO’ and ‘AC_VI’ delivery service.

The following example in Figure 5-11 explains the RREP processing at the destination and intermediate station. The white boxes at the source station ‘s1’ at two different time instances ‘t1’ and ‘t2’ show the arrival of different RREP messages for the same destination station. The source’s decision for selecting a route is based on the comparison, according to which the route having route metric $MADV_{RM2}$ will be selected if $MADV_{RM2} < MADV_{RM1}$.

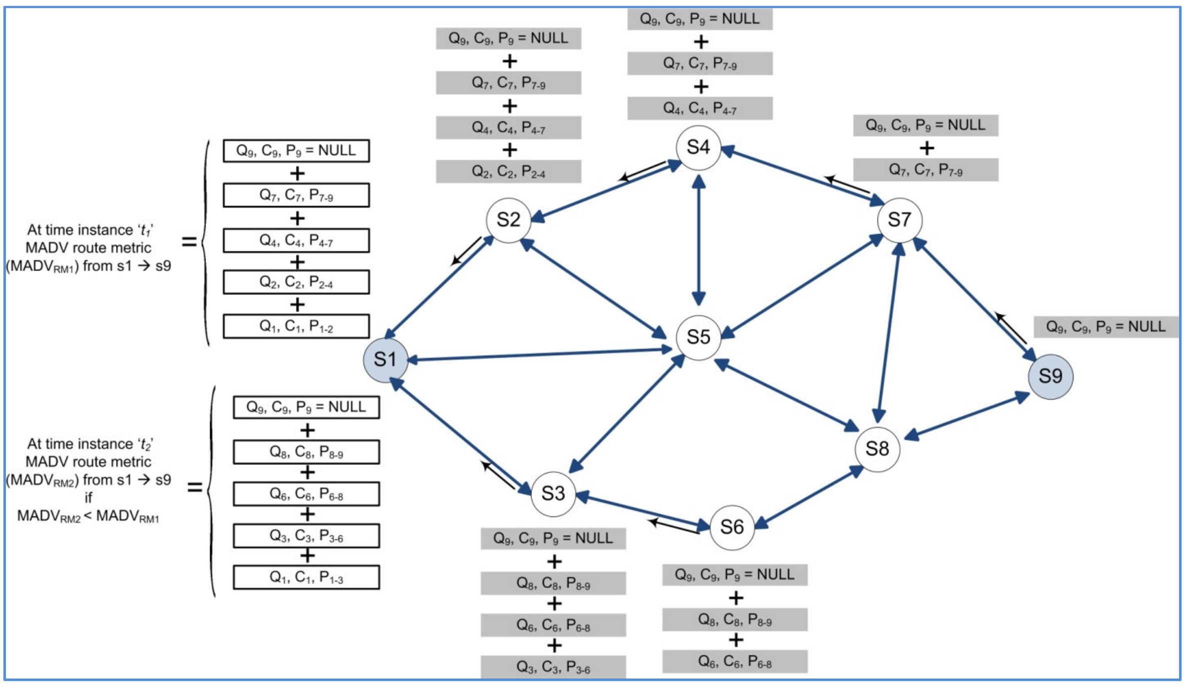

Figure 5-11: MADV RREP processing by a destination station and an intermediate station.

5.5.9 Periodic updates

In case of AODV local connectivity information is periodically refreshed by broadcasting HELLO packets. The HELLO message is essentially a RREP message with the exception that it is not unicast but broadcast, the destination sequence number field contains the latest sequence number of the station which broadcasts the HELLO packet, the TTL value is set to 1 and hop count is set to zero. The same mechanism of maintaining local connectivity is used in MADV.

In case of proactive routing protocols e.g. DSDV, routing updates propagate information about changes in the route entries of stations, by exchanging the routing table entries either by incremental updates or full dumps. If MADV is implemented in DSDV, the constituent parameters of MADV route metric which are all highly variable with time, can be better reflected through such updates and routing protocol can change routes with changes in the MAC dependent parameters.

As an example of MADV implementation in DSDV, the 'Stable data' field in the forwarding table should to an array of structures. Each structure would contain 'NextHop' and '$MADV_{RM}$' fields for every neighbour station. Elements of such an array would act as secondary routes. According to DSDV, routes which show improved metric are scheduled for advertisement at a time which depends on the average settling time for routes to a particular destination. Every station can update the MADV's QCP parameters periodically within the route update interval. With periodic updates, if there is a significant change in the MADV route metric of the primary and secondary routes, a route which offers the lowest metric value for a particular destination would be used in the 'primary' forwarding table. Table 5.4 shows the route entries for destination 's9' in the routing table of 's1', using DSDV, following the topology as shown in Figure 5-7. For the sake of simplicity fields like 'Sequence Number', 'Install', and 'flags' which are a part of a forwarding table in the original DSDV are not shown in this table. The four tables on the right show the alternative routes for each of the four ACs. In this case, for instance if the route where the next hop for AC_BK is 's3', offers a route metric lower than 7, then in the next update interval 's1' would swap the route entry for AC_BK, making 's3' as the next hop.

However, we believe that nevertheless the overall performance and the reaction of the routing protocol to changes at the MAC layer would be significantly optimized with this approach; it will undoubtedly increase the control overhead. As the MAC dependent parameters are highly variable, changes would trigger frequent route updates, and in the case of proactive routing mechanisms, such updates would be required to be disseminated across the network.

Therefore, we merely discussed the possibility of including MADV in a proactive routing protocol to provide an overview of the advantages and the disadvantages. However, our simulation models are only focused on the incorporation of MADV in a reactive (AODV) routing protocols owing to the factors as discussed above.

Table 5-4: Forwarding table of S1 (showing entry for destination station S9).

Destination	AC	NextHop	Metric ($MADV_{RM}$)	Stable-data
S9	AC_VO	S3	3.5	Ptr_S9_AC_VO
	AC_VI	S3	3.5	Ptr_S9_AC_VI
	AC_BE	S2	7	Ptr_S9_AC_BE
	AC_BK	S2	7	Ptr_S9_AC_BK

NextHop	Metric ($MADV_{RM}$)
s2	4
s5	4.5

NextHop	Metric ($MADV_{RM}$)
s2	4.5
s5	5

NextHop	Metric ($MADV_{RM}$)
s3	8
s5	8.5

NextHop	Metric ($MADV_{RM}$)
s3	8
s5	8.5

5.6 MADV's performance tests and comparative analysis

5.6.1 Simulation model

The MADV routing protocols has been modelled in OPNET, version 11.5A PL3. In this version of the OPNET network simulator a number of MANET routing protocols have been modelled including AODV, DSR and OLSR. As mentioned earlier that MADV is an enhancement of the AODV routing protocol, therefore, for implementation of the simulation model, the AODV model in OPNET has been modified by including the necessary changes as required by the MADV protocol.

In OPNET, a protocol's functionality is modelled in the form of an isolated (independent) model, referred to as a '*process model*'. Process models can maintain parent-child relationship, and depending on the model configuration a child process can be invoked by its parent process. A communication station's functionality is modelled by creating the OSI architecture within a station. At every layer of the OSI architecture there are separate modules, holding one or many process models (in case of parent-child relationship). A process model itself models the functionality of a protocol in the form of a state-transition diagram. The process can be in either of the available states. Changes/transition from one state to another state depends on certain conditions which are more related with the functionality of a protocol. For instance, a process in a '*Packet arrival*' state can be programmed to make a transition to another state (e.g. '*Process arrived packet*') when a packet arrives at one of the interface of the module holding this process model (this arrival is indicated by an interrupt) and so on. Various *functions* can be tied up with such transitions and the residing code within the function is executed each time such a transition is made from one state to another state. Figure 5-12 highlights the modelling hierarchy of the OPNET simulator.

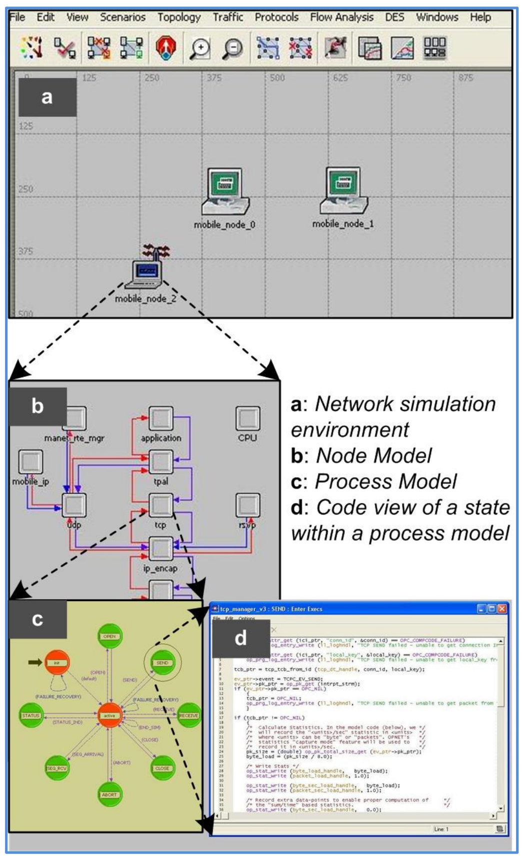

Figure 5-12: OPNET modelling hierarchy.

5.6.2 MADV model in OPNET

MADV is essentially based on the standard AODV protocol. It follows the same procedure for *route-discovery, route maintenance, and route updates.* However, MADV does not use the *hop-count* as a route selection metric during the *route-request* and *route-reply* phases of the route discovery process, but instead it uses the queuing delays experienced by frames in all four access category queues (i.e. AC_VO, AC_VI, AC_BE and AC_BK), the contention delay experienced by the MAC sub-layer and the propagation time that a frame would take when transmitted by the transmitter to a particular receiver pair. All this information is MAC and in turns PHY specific. For the realization of MADV this information has to be made available to the MADV routing demon/process model. This requirement makes MADV a classical *cross-layer design* scenario, where the decision of a higher layer protocol is dependent on the information of the lower layers.

In OPNET, the MAC level information is passed to the higher layers by using an inter-process communication method of setting attributes by the process at the lower layers and then reading the attributes by any relevant processes at the higher layers.

5.6.2.1 Reporting Contention delay

The contention delay is calculated at the MAC layer with in the '*wlan_mac_hcf'* process model. When a frame is sent to the transmit buffer, the *functions* which are called for dealing with the transmission process, stores the time of arrival of such frames, this is stored in '*hld_ptr->time_rcvd*'. At the time when these function determine that medium is idle and the frame is ready for transmission, it simply subtracts the '*hld_ptr->time_rcvd*' from the '*current_time*' variable to get the contention delay experienced by this frame. The value of the contention delay recorded at the MAC layer is then written as a value of an attribute (which in this case is called *'MADV_Cd'*, and is declared as an attribute at the 'ip' process model).

5.6.2.2 Reporting propagation delay

The propagation delay is the time taken by sending a frame at a particular transmission rate. This value varies with variation in the frame size and depends on the transmission rate. The propagation delay, denoted by P_d is calculated by dividing the size of

MSDU with the current transmission rate[11]. The calculated value of P_d is then communicated to the IP layer.

5.6.2.3 Reporting queuing delay

At the MAC sublayer there are four different queues which handle the arrival of frames with four different access categories. At one time the ‘wlan_mac_hcf’ takes only one frame (after resolution of the inter-ACs contention) and handles it transmission; the access category is identified from the value in the variable *‘cur_tx_ac’*. Based on the type of AC, the corresponding queuing delay is reported to the higher layers[12].

5.6.3 Simulation environment

The simulation environment consists of 21 wireless stations (station model: wlan_wkstn_adv) constituting a wireless ad-hoc network. The IP addressing scheme is IPv4. The PHY is configured to be conformant with the IEEE 802.11g (Extended rate PHY). The MAC is enabled to be conformant with the IEEE 802.11e and default EDCA parameters are used in the simulation.

The source station (IP address of 192.168.1.2) is configured to support different profiles in different simulation scenarios, where each profile is configured to have different applications. Four custom applications were configured: (1) custom_app_AC_BK, (2) custom_app_AC_BE, (3) custom_app_AC_VI and (4) custom_app_AC_VO. The four applications generate outbound traffic at the same rate, but the ‘Type of Service’ is different for each of the four applications, for example the configuration for custom_app_AC_BK is given below in Table 5-5.

---

[11] MADV_propagation_time = data_size/operational_speed;
MADV_node_id = op_topo_parent (op_idself());
MADV_int_inf_Objid=op_id_from_name(MADV_node_id, OPC_OBJTYPE_PROC, “ip”);
Op_ima_obj_attr_get(MADV_int_inf_objid, “MADV_Pd”, MADV_propagation_time);

[12] MADV_Qd_AC[cur_tx_ac][0]=current_time – hld_ptr->time_rcvd;
If(cur_tx_ac==0)
op_ima_obj_attr_set(MADV_int_inf_Objid, “MADV_Qd_AC_BK”, MADV_Qd_AC[cur_tx_ac][0];
…
If(cur_tx_ac==3)
op_ima_obj_attr_set(MADV_int_inf_Objid, “MADV_Qd_AC_VO”, MADV_Qd_AC[cur_tx_ac][0];

Table 5-5: Application configuration parameters for custom_app_AC_BK application (low resolution video).

Attribute	Value
Frame Inter-arrival Time Information	10 frames/sec
Frame Size Information	128 x 120 pixels
Symbolic Destination Name	Video Destination
Type of Service	**Background**
Traffic Mix	All Discrete

For the other three applications, the 'Type of Service' field is changed to 'Best Effort', 'Streaming Multimedia' and 'Interactive Voice' respectively. Therefore, each of the four applications generates 3686400 bits/sec[13]. This traffic configuration is by default called 'Low resolution video' in the OPNET application definition manager.

The stations in the ad-hoc network are positioned as shown in the Figure 5-13. Here in this scenario, stations with IP addresses from 192.168.1.13-15 and 192.168.1.22 uses transmission rate of 1 Mbps, while all other stations operate at 54 Mbps. From this scenario, there are three possible routes between the source and the destination stations:

1. **Route-a**: consists of 10 hops between the source and the destination stations. All the stations along the route-a operate at 54 Mbps.
2. **Route-b**: consists of 8 hops between the source and the destination stations. Stations with IP addresses from 192.168.1.13-15 and 192.168.1.22 uses transmission rate of 1 Mbps, while all other stations operate at 54 Mbps.
3. **Route-c:** consists of 10 hops between the source and the destination stations. Station with an IP address of 192.168.1.22 which is also a part of route-b operates at 1 Mbps while all other stations along this route operate at 54 Mbps.

The scenario and application configuration as given in Figure 5-13 and Table 5-5 are configured for achieving the following objectives:

1. To demonstrate the ability of a routing protocol (in this case the MADV) to perform route selection by including MAC level parameters.
2. Show the effect of baseline traffic on the route selection process.

[13] 128 x 120 = 15360 pixels in each frame. Each pixel is represented by 3 bytes which implies that each pixel's information is represented by 3 x 8=24 bits. Therefore, total number of bits in each frame of the video application = 128 x 120 x 3 x 8=368640 bits/frame.
The total traffic generated by the video application in terms of bits/sec= frame size * frame rate per sec = 368640 x 10 =3686400 bits/sec.

3. Show the effect of baseline traffic on network traffic belonging to various access categories and how a routing protocol can optimize the route selection process in the wake of different baseline traffic patterns on all available routes.
4. To demonstrate in further depth, the effect of transmission rates, queuing delays and contention experienced at the MAC layer on the overall route performance in terms of delay and end to end throughput.
5. To demonstrate the relationship that the three MAC level parameters (i.e. the transmission rate, queuing and contention delay) have and the effect of such relationship on the overall route selection.

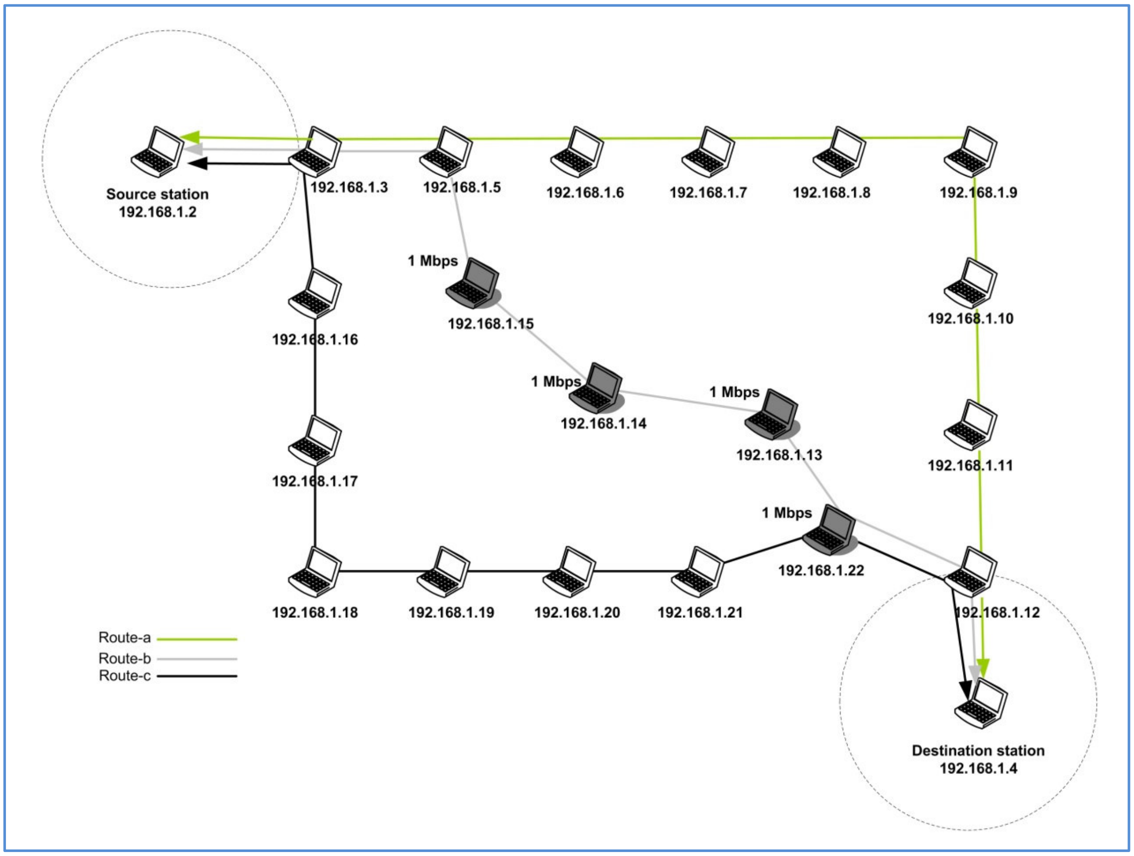

Figure 5-13: Simulation scenario showing three possible routes between 'source' and 'destination' stations.

5.6.4 Comparative analysis

In order to show the performance of the MADV in comparison with AODV and DSR in the scenario given in Figure 5-13, we initially configured scenarios for measuring the performance of the three possible routes with various application profiles on the source station. With the performance of each of the three possible routes at hand, we can better evaluate the performance of a routing protocol.

In the first simulation run, three scenarios were configured; scenario-1 used route-a between the source and destination station, scenario-2 used route-b and scenario-3 used route-c. The traffic configuration on the source station remained the same except that on the scenario-2, the 'Type of service' for the traffic is 'Streaming Video' while that for the other two scenarios is 'Interactive Voice'.

Figure 5-14 shows the end-to-end delay experienced by packets from source to destination station along the route-a. The figure shows two curves one for AC_BE, which shows the routing protocol's control traffic and AC_VI which shows the baseline traffic. It is important to notice that the values of end-to-end delay shown in Figure 5-14 and the following figures showing end-to-end delay are sum of the queuing, contention and propagation delays experienced by a frame hop-by-hop as it travels along the route.

It is obvious from Figure 5-14 that the contention delay in this scenario is almost negligible. Likewise, all the stations along the route-a operate at 54 Mbps, which makes the propagation delay to a minimal importance in comparison with the queuing delays. In essence it is true to infer that the delay values shown in Figure 5-14 are significantly the queuing delays, therefore, prioritization of ACs would seriously affect the performance of some ACs (such as the BK and BE). This argument is revisited later in the following discussion.

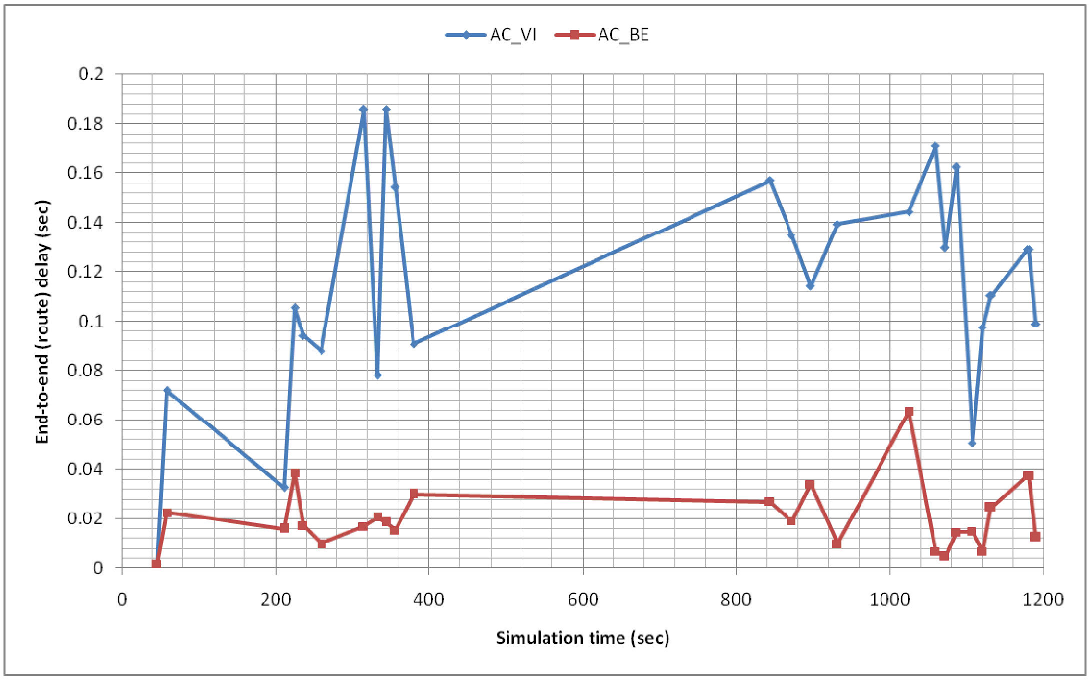

Figure 5-14: End-to-end (route) delay of frames in access categories *Video* and *Best Effort* between source and destination stations along route-a, showing the baseline traffic.

For demonstrating the effect of queue prioritization in a scenario where the end-to-end delay is significantly affected by queuing delay (which is the case of route-a), we configured four different scenarios. In each of the four scenarios, we configured another application where the traffic delivery service is specified to be in a different AC along with the baseline traffic load on the source station.

Scenario-route-a-1: Has the baseline traffic and another application which generates traffic at the same rate but the traffic delivery service is set to 'Background'. Therefore, in this scenario the network packets would fall in either of the three ACs (1) AC VO, for the baseline traffic, (2) AC BE for the routing control traffic, which is dependent on the routing protocol and (3) AC BK, for the traffic added to the baseline load in this scenario.

Scenario-route-a-2: Baseline traffic of route-a, and another application where the delivery service is set to 'Best Effort'. In this scenario the network packets would fall in any of the two ACs (1) AC VO, for the baseline traffic, (2) AC BE for the routing control traffic and for the traffic added to the baseline load in this scenario.

Scenario-route-a-3: Baseline traffic of route-a, and another application where the delivery service is set to 'Streaming Video'. In this scenario the network packets would fall in any of the two ACs (1) AC VO, for the baseline traffic and for the traffic added to the baseline load in this scenario, and (2) AC BE for the routing control traffic.

Scenario-route-a-4: Baseline traffic of route-a, and another application where the delivery service is set to 'Interactive Voice'. In this scenario the network packets would fall in any of the three ACs: (1) AC VI, for the traffic added to the baseline load in this scenario, (2) AC VO, for the baseline traffic and (2) AC BE for the routing control traffic.

Figure 5-15 shows the end-to-end delay experienced by frames belonging to AC-BK, AC-BE and AC-VI along the route-a. The rise and fall in the delay curves is due to the configuration of the application profile, where the applications run for a particular duration of time, and then the applications are restarted after a particular inter-repetition interval. From Figure 5-15, it is evident that traffic belonging to AC-BK on route-a would experience almost transmission disruption (and eventually a very higher delay) when a similar traffic load with a queue priority (AC-VI) is simultaneously active on the same route.

Results from 'Scenario-route-a-2', 'Scenario-route-a-3', and 'Scenario-route-a-4' are shown in Figure 5-16, Figure 5-17 and Figure 5-18 respectively. It is evident from the results that in the presence of the baseline load (with higher priority), lower priority traffic would suffer higher end-to-end delays mainly due to queuing.

These figures give an interesting insight into the queue prioritization when the traffic load falls into different ACs. According to the IEEE 802.11e standard specification, a higher priority AC wins in the case of internal collision. Therefore, when traffic load increases in AC-queues of higher priority it affects the lower priority queues frames significantly.

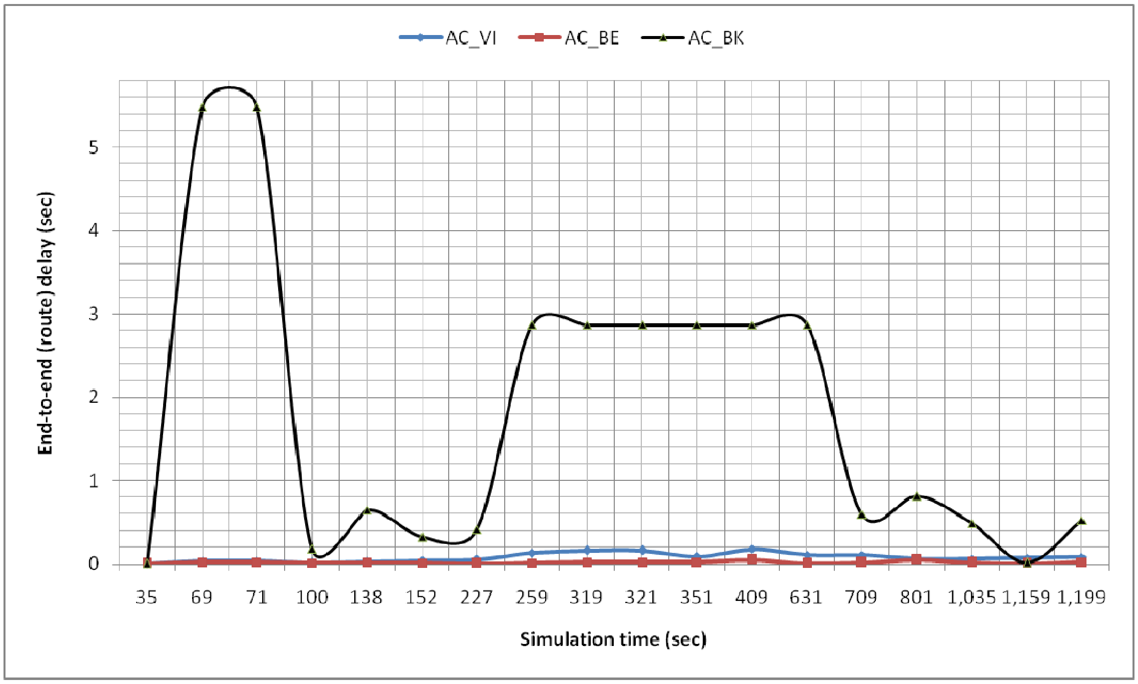

Figure 5-15: End-to-end (route) delay of frames in access categories *Video*, *Best Effort* and *Background* between source and destination stations along route-a. This plot is from scenario-route-a-1.

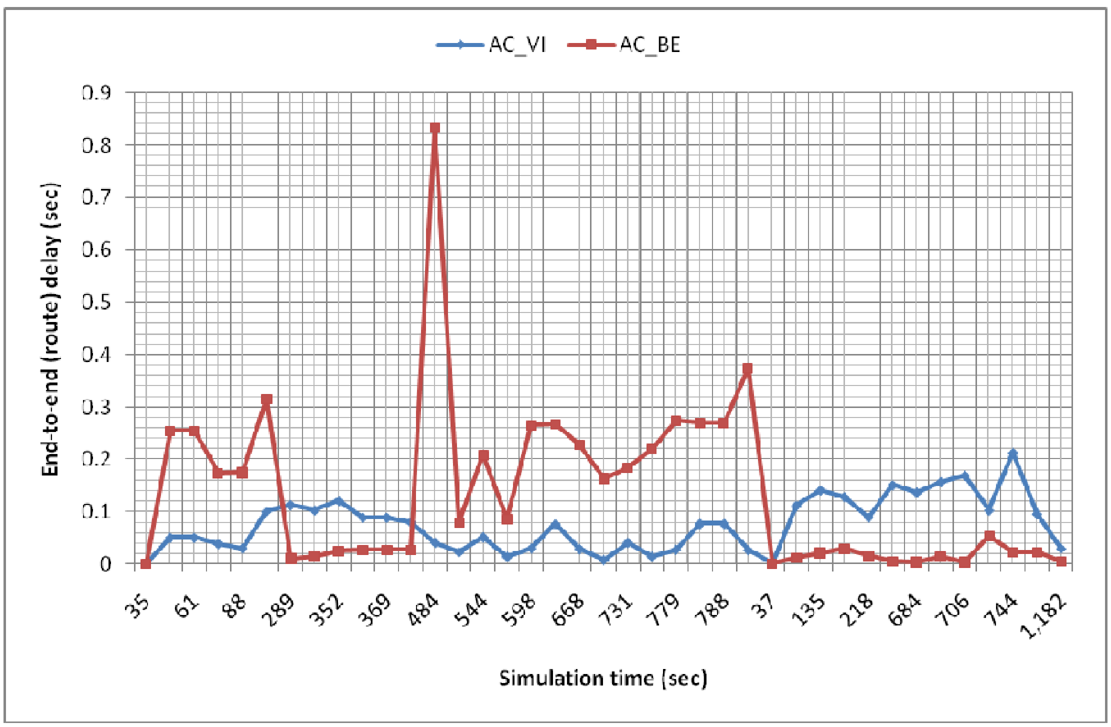

Figure 5-16: End-to-end (route) delay of frames in access categories Video and Best Effort between source and destination stations along route-a. This plot is from scenario-route-a-2.

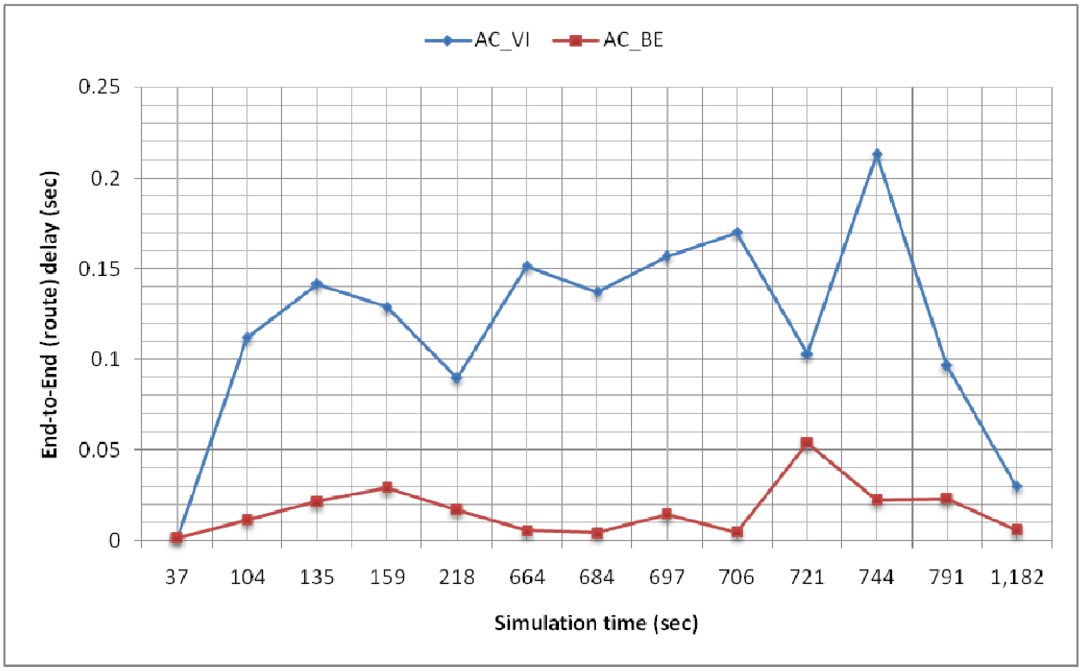

Figure 5-17: End-to-end (route) delay of frames in access categories Video and Best Effort between source and destination stations along route-a. This plot is from scenario-route-a-3.

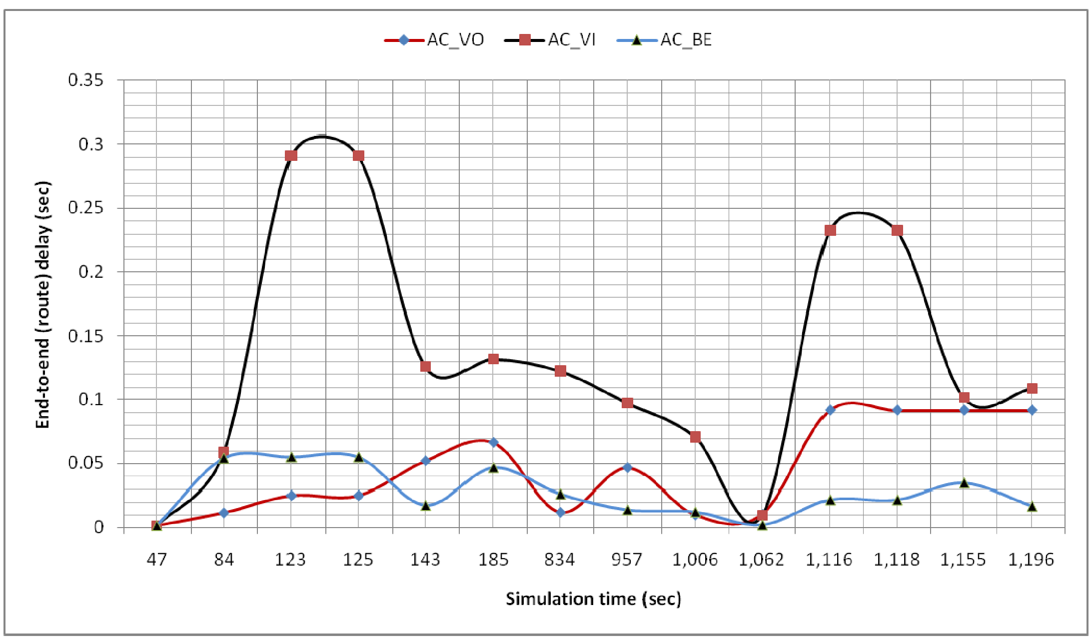

Figure 5-18: End-to-end (route) delay of frames in access categories *Voice, Video* and *Best Effort* between source and destination stations along route-a. This plot is from scenario-route-a-4.

In order to demonstrate the route characteristics of route-b and route-c, we configured two more scenarios with baseline traffic similar to that of the route-a, but the delivery service of the baseline traffic is set to 'Interactive Voice'. Figure 5-19 and Figure 5-20 show the end-to-end delay experienced along the route-b and route-c respectively by frames belonging to the AC-VO and AC-BE. It is clearly evident that the end-to-end delay experienced by the baseline traffic is significantly higher in the case of route-b when compared to route-a; in the case of route-a, the delay values roughly lie between 0.03 to 0.19 seconds while that in the case of route-b lie roughly between 0.5 to 2.5 seconds with an average value just close to 1 second. The end-to-end delay of the baseline traffic on route-c has an average value of around 0.5 second.

It is important to consider that here in this situation the rate at which the source generates the traffic is higher than the transmission rate of some stations along the route-b. Stations with lower transmission rate become a bottle neck in the overall throughput, increasing the queuing delay because the arrival rate of the packets in the queue is higher than the removal rate. A comparison of the queuing delay experienced by frames of AC-VO at station with IP address 192.168.1.21 and 192.168.1.22 (which is a common station of route-b and route-c with a transmission rate of 1 Mbps) is shown in Figure 5-21.

Changing the source traffic generation rate to a lower value (less than the transmission rate of the lowest transmission rate of any station along the route) would significantly decrease the queuing delays. In such situation the number of hops would then play an important role, because when the queuing delay is of minimal importance, the propagation time (transmission rate) plays a significant role in the end-to-end delay of route. Figure 5-22 shows the end-to-end delay of route-b when the stations which previously used transmission rate of 1 Mbps have been set to use 11 Mbps. The queuing delay is scaled down to 1/5th of the delay values as reported in Figure 5-19.

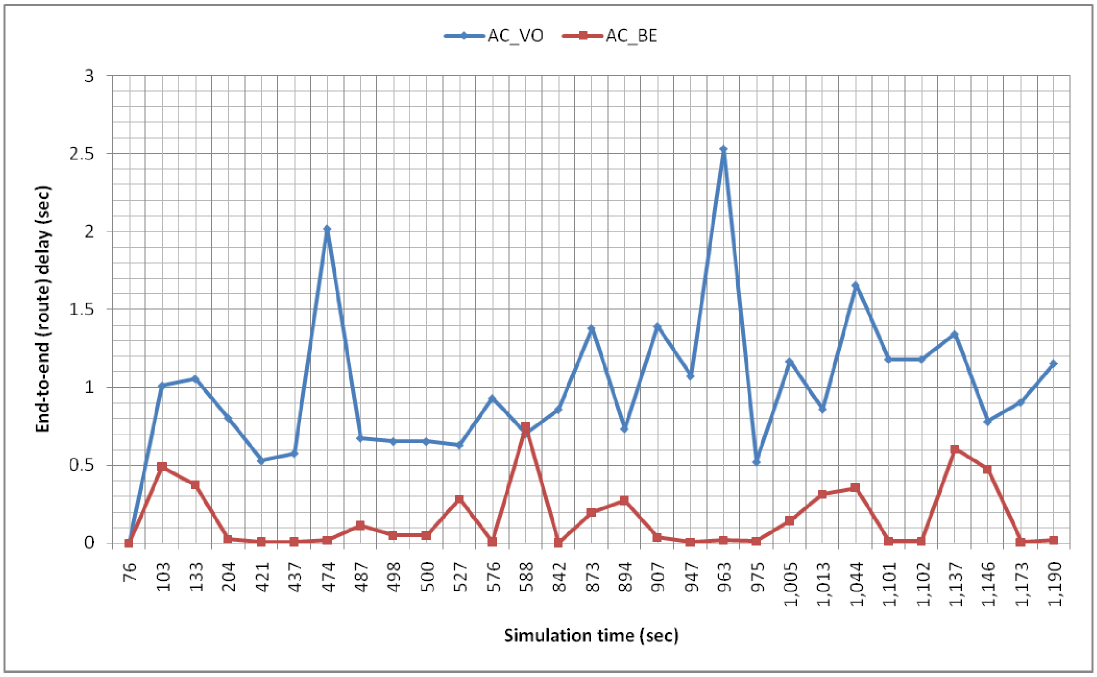

Figure 5-19: End-to-end (route) delay of frames in access categories *Voice* and *Best Effort* between source and destination stations along route-b.

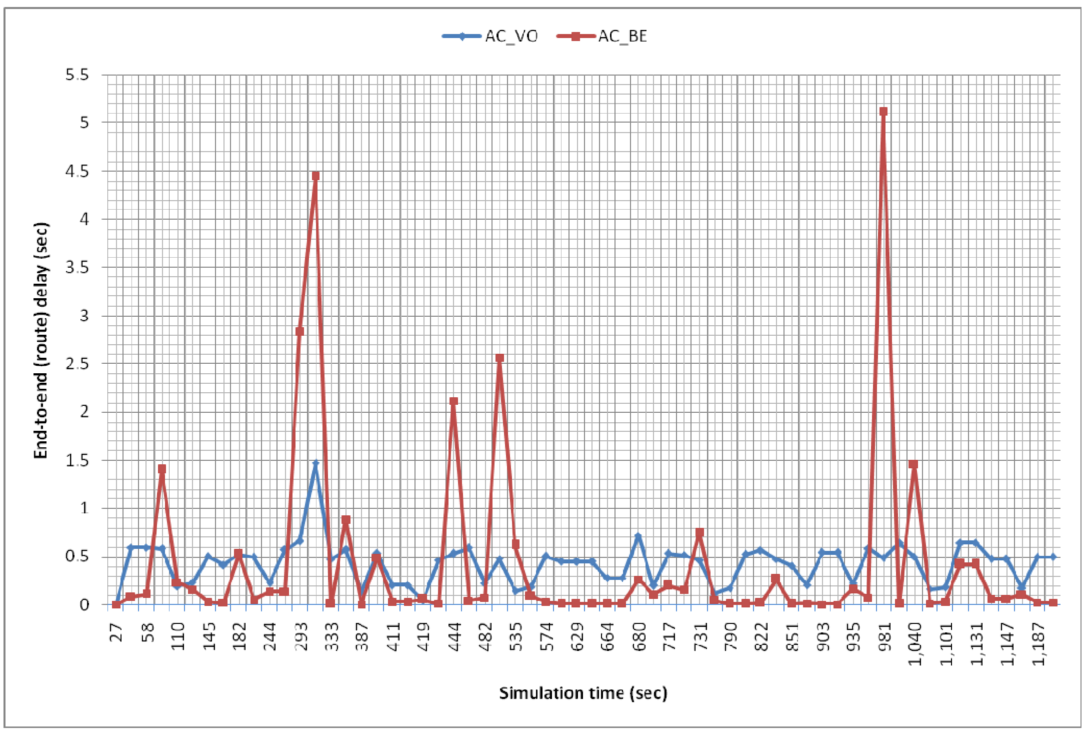

Figure 5-20: End-to-end (route) delay of frames in access categories *Voice* and *Best Effort* between source and destination stations along route-c.

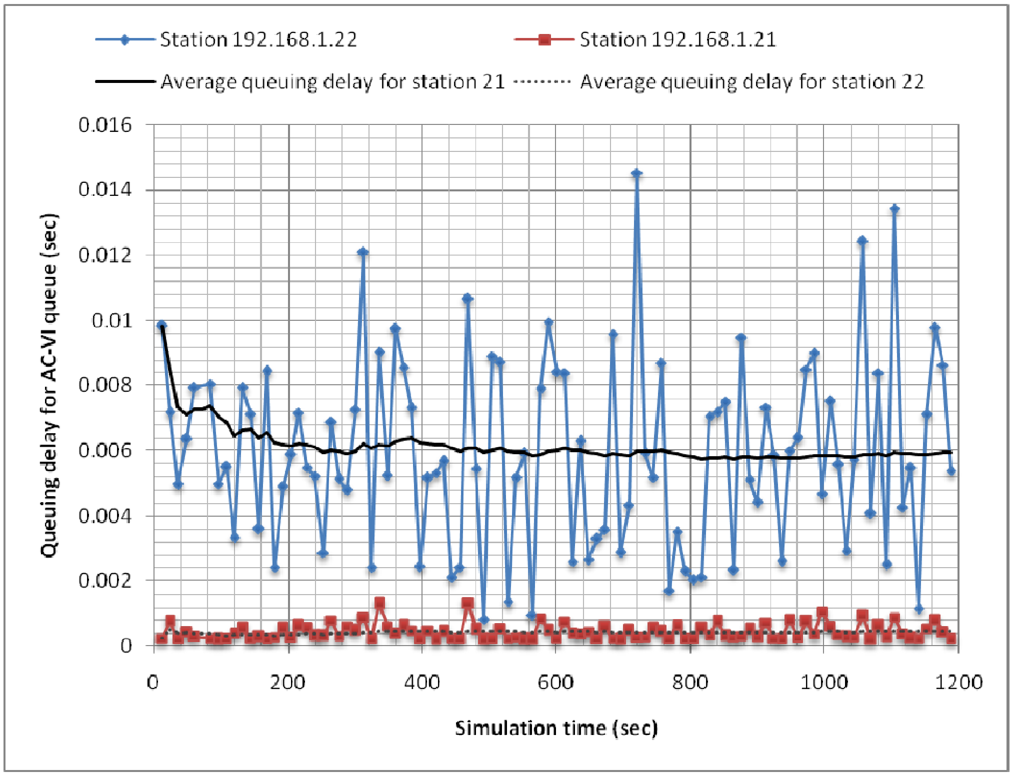

Figure 5-21: Comparison of queuing delay experienced by frames belonging to AC-VO at stations with IP addresses: 192.168.1.21 (operating at 54 Mbps) and 192.168.1.22 (1 Mbps).

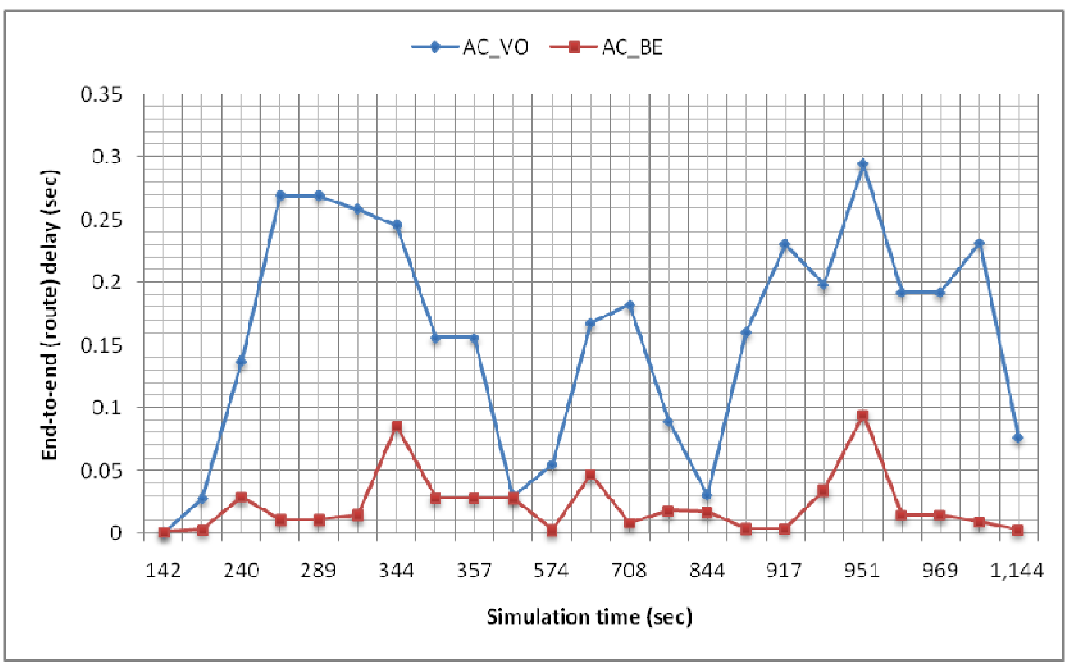

Figure 5-22: End-to-end (route) delay of frames in access categories Voice and Best Effort between source and destination stations along route-b. The stations which operated at transmission rate of 1 Mbps are set to 11 Mbps in this scenario.

So far, the previous discussion has attempted to highlight the significance of the underlying MAC level parameters on the overall performance of the end-to-end route metrics. MADV is designed to take into consideration these important MAC level parameters and take an intelligent routing decision which is most suitable for an AC type under the prevailing network conditions. It is important to notice that MADV does not maintain a single route for all types of ACs, (refer to section 5.5.7 above), this feature is important in some scenarios. For instance consider the topology shown in the Figure 5-13, and consider that the route-b has baseline traffic in the AC-BK and the source station has to route traffic in the AC-BK. The route-a has baseline traffic in the AC-VI. In such scenario, it is highly likely that using route-b for the AC-BK traffic would offer less delay than routing it through the route-a (because of the presence of a higher priority AC already active). In this scenario, the station with IP address 192.168.1.5 would forward the AC-BK packets to 192.168.1.15.

As a further step in the performance testing, we configured the stations in the topology shown in Figure 5-13 to run AODV and in another scenario to run DSR as a routing protocols and the source station was configured to have the baseline load. As mentioned earlier that AODV and DSR uses a shortest path metric for route selection, therefore, both AODV and DSR selects route-b for routing traffic between the source and the destination station.

Figure 5-23 shows the end-to-end throughput of routes selected by MADV, AODV and DSR between the source and the destination station as given in the network topology of Figure 5-13. The drop and rise shown by the throughput plot is because of the application configuration, the drop areas reflect the time when the application did not generate any network traffic. Figure 5-24 shows the time average plot of the end-to-end throughputs for MADV, AODV and DSR.

This obviously is the worst case scenario where the AODV and DSR which inherently relies on the hop-count are completely mislead in making an intelligent decision of route selection according to the underlying MAC level parameters. The throughput of AODV and DSR is evidently unmatched with the throughput of the route selected by MADV. Addition of another traffic stream of a different AC between the same source and destination would present an interesting comparison scenario. AODV and DSR would simply lookup their routing tables and then forward it along the same route as for the first stream, which would overload the same route, causing prolonged queuing delays and an overall longer end-to-end delivery time. While MADV on the other hand would select a route based on the most appropriate metric values per AC at any given time.

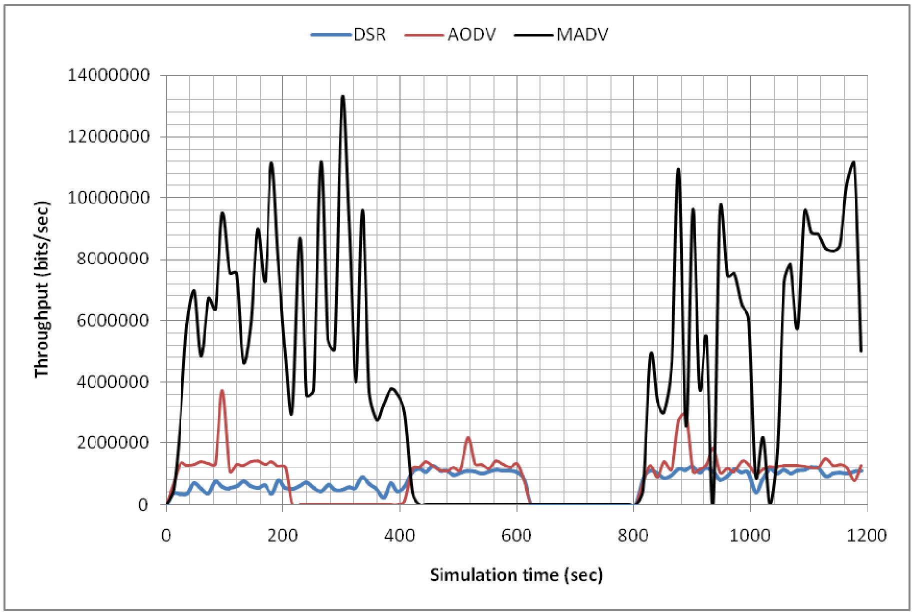

Figure 5-23: Throughput comparison of MADV with DSR and AODV using baseline traffic loads. The route used by the MADV has baseline traffic configured to use the AC_VI delivery service while the route used by the DSR and AODV has baseline configured in AC_VO.

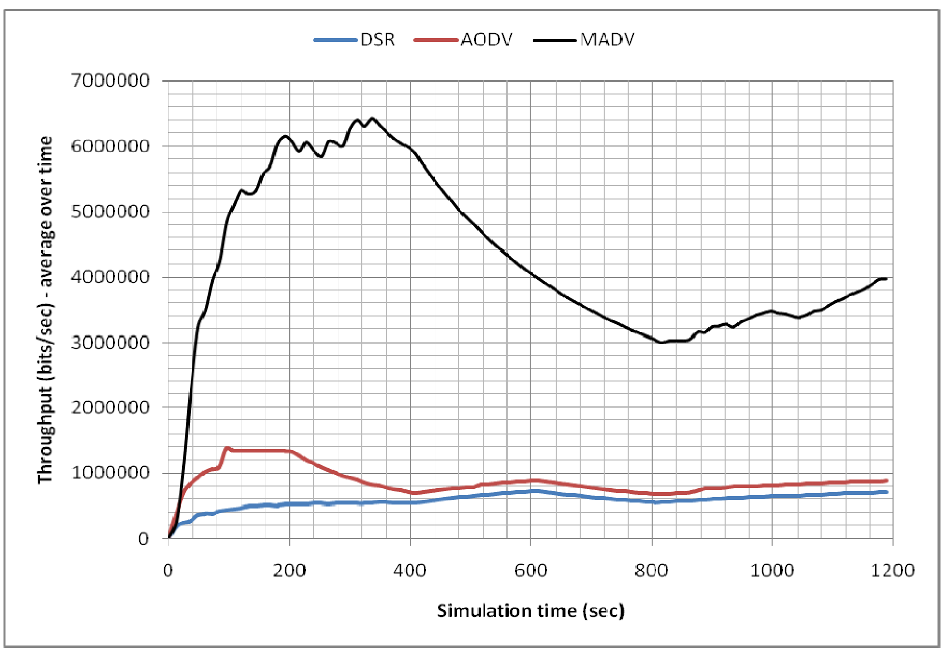

Figure 5-24: Throughput comparison of MADV with DSR and AODV using baseline traffic loads. The result is plotted as a time-average.

5.7 Conclusion

This chapter presented an in depth analysis of the effect of the MAC level parameters on the performance of routing protocols. The MAC sub-layer is only responsible for handling single hop communication within a BSS area (where stations operate using a single/same coordination function). On the other hand a routing protocol is responsible for end-to-end route establishment and delivery of packets to destination. However, there are several MAC level factors such as the transmission rate, medium access contention and queuing delays, which are variable with time and have direct or indirect influence on the performance of a routing protocol. The IEEE 802.11e introduces four different MAC level queues for each of the four access categories. The standard specifies maintenance of service priority within the queues; which implies that frames from a higher priority queue would be serviced more frequently than frames belonging to lower priority queues. Such an enhancement at the MAC sub-layer introduces uneven queuing delays. Conventional/existing routing protocols are unaware of such queue prioritization and as a result these factors are not considered which cause severe performance deterioration for frames belonging to lower priority queues. It is important to consider hop-by-hop (or a link) MAC level information while constructing an end-to-end routing strategy for a multi hop wireless network.

Towards this end we formulated a routing metric and incorporated it into the Medium Aware Routing Protocol (MADV), which combines variations in the transmission rates, contentions levels and queuing delay at every link. The routing table structure is modified from the existing routing protocol's approaches by introducing per destination-per AC entry; which essentially means that a source station can have different route entries based on the ACs for the same destination. The advantage of ACs-specific route entries is that MAC-level queue prioritization can be effectively taken into account for the whole route, and the route selection is performed based on the lowest metric value per AC from the available routes to a destination.

MADV decouples the conventional notion of hop-count from the underlying network performance. Parameters like the queuing and contention delay are dependent on the selection of appropriate *stations* along the route which offer minimum queuing and contention delays. On the other hand, minimizing the propagation delay requires the selection of *links* which offer minimum propagation delay. In both cases, the 'number of hops' metric is completely independent from the parameters which affect the overall communication and thus has no role in improving performance in multi-hop communication. Therefore, it is important to decouple the hop-count metric which is conventionally associated to best path, by integrating the MAC layer parameters into the routing protocol.

MADV is an alternative approach of using the underlying MAC-layer parameters in construction of the routing metric and our objective with the use of MADV is not to implement yet another routing protocol for routing in mobile ad-hoc multi-hop networks, but to integrate the MAC layer enhancements into the routing protocols in the form of the MADV metric for performance enhancements. As discussed in the 5.3, standalone MAC layer enhancements when not taken into consideration can adversely affect the performance of routing protocols. MAC layer is the closest level to monitor the channel state information of a time-varying quality of channels. We believe that by integrating the MAC layer parameters, which are significantly important for determining the quality of communication in terms of various QoS parameters (e.g. delay, throughput, frame error rate etc), we can significantly improve the performance of existing routing protocols.

The MADV uses the same approach as AODV for route discovery and maintenance, with slight modification for integrating the new route metric for route selection. Route discovery is performed with the help of route request (RREQ) and route reply (RREP) initiated by the source and destination/intermediate stations respectively.

The MADV-metric can be incorporated into various routing protocols and its applicability is determined by the possibility of provision of MAC dependent Q.C.P parameters which are used to determine the hop by hop MADV-metric values.

5.8 References

[1] J. Jubin and J. D. Tornow, The DARPA packet radio network protocols. Proceedings of the IEEE, 75(1): 21-32, January 1987.

[2] G. S. Lauer, Packet-radio routing. In Routing in Communications Networks, edited by Martha E. SteenStrup, Prentice-Hall, Englewood Cliffs, New Jersey, 1995.

[3] X. Chen, H. M. Jones and A. D. S. Jayalath, "Congestion-Aware Routing Protocol for Mobile Ad Hoc Networks," ,Vehicular Technology Conference, 2007. VTC-2007 Fall. 2007 IEEE 66th 21-25, 2007.

[4] S. Shakkottai, T. Rappaport and P. Karlsson, "Cross-layer design for wireless networks," Communications Magazine, IEEE, vol. 41, pp. 74-80, 2003.

[5] V. Srivastava and M. Motani, "Cross-layer design: a survey and the road ahead," Communications Magazine, IEEE, vol. 43, pp. 112-119, 2005.

[6] M. van Der Schaar, N. Sai Shankar, "Cross-layer wireless multimedia transmission: challenges, principles, and new paradigms," Wireless Communications, IEEE , vol.12, no.4, pp. 50-58, Aug. 2005

[7] N. Nang, R. Sankar and Jungsik Lee, "Improving ad hoc network performance using cross-layer information processing," pp. 2764-2768, 2005.

[8] V. Kawadia and P. Kumar, "A cautionary perspective on cross-layer design," Wireless Communications, IEEE [See also IEEE Personal Communications], vol. 12, pp. 3-11, 2005.

[9] C. E. Perkins and P. Bhagwat, "Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers," in 1994, pp. 234-244.

[10] T. Clausen , P. Jacquet, Optimized Link State Routing Protocol (OLSR), RFC Editor, 2003.

[11] Ogier, R., Templin, F., and Lewis, M. 2004 Topology Dissemination Based on Reverse-Path Forwarding (Tbrpf). RFC. RFC Editor.

[12] G. Pei, M. Gerla, and T.-W. Chen, “Fisheye State Routing: A Routing Scheme for Ad Hoc Wireless Networks,” in Proc. ICC 2000 New Orleans, LA, June 2000.

[13] C. E. Perkins, E. Belding-Royer, S.R. Das. Ad hoc on-demand distance vector (AODV) routing. http://www.ietf.org/rfc/rfc3561.txt , July 2003. RFC 3561.

[14] D. B. Johnson,, D. A. Maltz, , and J. Broch, 2001. DSR: the dynamic source routing protocol for multihop wireless ad hoc networks. In Ad Hoc Networking Addison-Wesley Longman Publishing Co., Boston, MA, 139-172.

[15] V. Park, S. Corson, Temporally-Ordered Routing Algorithm (TORA) Version 1Functional Specification, IETF MANET Working Group INTERNET-DRAFT. http://tools.ietf.org/id/draft-ietf-manet-tora-spec-04.txt

[16] Z. Haas and M. Perlman, The Zone Routing Protocol (ZRP) for ad hoc networks, Internet draft, Mobile Ad-Hoc Network (MANET) Working Group, IETF (1998).

[17] Haas, Z.J., Pearlman, M.R. and Samar, P., "Intrazone Routing Protocol (IARP)," IETF Internet Draft, draft-ietf-manet-iarp-02.txt, July 2002.

[18] Haas, Z.J., Pearlman, M.R. and Samar, P., "Interzone Routing Protocol (IERP)," IETF Internet Draft, draft-ietf-manet-ierp-02.txt, July 2002.

[19] S. Basagni, , I. Chlamtac, V. R. Syrotiuk, and B. A. Woodward, 1998. A distance routing effect algorithm for mobility (DREAM). In Proceedings of the 4th Annual ACM/IEEE international Conference on Mobile Computing and Networking (Dallas, Texas, United States, October 25 - 30, 1998). W. P. Osborne and D. Moghe, Eds. MobiCom '98. ACM, New York, NY, 76-84.

[20] B. Karp , H. T. Kung, GPSR: greedy perimeter stateless routing for wireless networks, Proceedings of the 6th annual international conference on Mobile computing and networking, p.243-254, August 06-11, 2000, Boston, Massachusetts, USA.

[21] C.-C. Chiang, H-K Wu, W. Liu, and M. Gerla, “Routing in clustered multihop, mobile wireless networks,” in Proc. IEEE Singapore Int. Conf. Networks, 1997, pp. 197–211.

[22] M. Gerla and J. Tsai, “Multicluster, mobile, multimedia radio network,” ACM-Baltzer J. Wireless Networks, vol. 1, no. 3, pp. 255–265, 1995.

[23] A. Iwata, C.-C. Chiang, G. Pei, M. Gerla, and T.-W. Chen, Scalable routing strategies for ad hoc wireless networks. IEEE Journal on Selected Areas in Communications, Special Issue on Ad-Hoc Networks, August 1999, p1369-p1379.

[24] E. M. Belding-Roye. 2003. Multi-level hierarchies for scalable ad hoc routing. Wirel. Netw. 9, 5 (Sep. 2003), 461-478.

[25] Mingliang Jiang, Jinyang Li and Y. C. Tay. Cluster Based Routing Protocol (CBRP), Internet draft,draft-ietf-manet-cbrp-spec-01.txt.

[26] C. C. Chiang, T. C. Tsai, W. Liu and M. Gerla, Routing in clustered multihop, mobile wireless networks with fading channel, The Next Millennium, The IEEE SICON, 1997.

[27] P. Johansson , T. Larsson, N. Hedman, B. Mielczarek, and M. Degermark. 1999. Scenario-based performance analysis of routing protocols for mobile ad-hoc networks. In Proceedings of the 5th Annual ACM/IEEE international Conference on Mobile Computing and Networking (Seattle, Washington, United States, August 15 - 19, 1999). MobiCom '99. ACM, New York, NY, 195-206

[28] J. Broch, D. A. Maltz, D.B Johnson, Y. Hu, and J. Jetcheva. 1998. A performance comparison of multi-hop wireless ad hoc network routing protocols. In Proceedings of the 4th Annual ACM/IEEE international Conference on Mobile Computing and Networking (Dallas, Texas, United States, October 25 - 30, 1998).

[29] H. Ehsan, Z. A. Uzmi, "Performance comparison of ad hoc wireless network routing protocols," Multitopic Conference, 2004. Proceedings of INMIC 2004. 8th International , vol., no., pp. 457-465, 24-26 Dec. 2004.

[30] J. Raju, J. J, Garcia-Luna-Aceves, "A comparison of on-demand and table driven routing for ad-hoc wireless networks," Communications, 2000. ICC 2000. 2000 IEEE International Conference on , vol.3, no., pp.1702-1706 vol.3, 2000.

[31] A. Boukerche, Performance evaluation of routing protocols for ad hoc wireless networks. Mob. Netw. Appl. 9, 4 (Aug. 2004), 333-342.

[32] L. Hui , Y. Dan, "Comparison of ad hoc and centralized multihop routing," Wireless Personal Multimedia Communications, 2002. The 5th International Symposium on , vol.2, no., pp. 791-795 vol.2, 27-30 Oct. 2002.

[33] E. M. Royer, T. Chai-Keong, "A review of current routing protocols for ad hoc mobile wireless networks," Personal Communications, IEEE , vol.6, no.2, pp.46-55, Apr 1999.

[34] Y. Tao, M. Ikeda, G. De Marco, L. Barolli, "Performance Behavior of AODV, DSR and DSDV Protocols for Different Radio Models in Ad-Hoc Sensor Networks," Parallel Processing Workshops, 2007. ICPPW 2007. International Conference on , vol., no., pp.6-6, 10-14 Sept. 2007.

[35] "IEEE Standard for Information technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements

- Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," ,IEEE Std 802.11-2007 (Revision of IEEE Std 802.11-1999) C1-1184, 2007.

[36] S. Zhao, Z. Wu, A. Acharya and D. Raychaudhuri, "PARMA: A PHY/MAC aware routing metric for ad-hoc wireless networks with multi-rate radios," in 2005, pp. 286-292.

[37] H. Lundgren, E. Nordströ and C. Tschudin, "Coping with communication gray zones in IEEE 802.11 b based ad hoc networks," in 2002, pp. 49-55.

[38] A. Kamerman and L. Monteban, "WaveLAN®-II: a high-performance wireless LAN for the unlicensed band," Bell Labs Technical Journal, vol. 10, pp. 119, 1997.

[39] S. Khan, S. A. Mahmud, K. K. Loo and H. S. Al-Raweshidy, "A cross layer rate adaptation solution for IEEE 802.11 networks," Computer Communications, vol. 31, pp. 1638-1652, 5/25. 2008.

[40] M. Lacage, M. H. Manshaei and T. Turletti, "IEEE 802.11 rate adaptation: a practical approach," Proceedings of the 7th ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems, pp. 126-134, 2004.

[41] Z. Li, A. Das, A. Gupta and S. Nandi, "Full auto rate MAC protocol for wireless ad hoc networks," Communications, IEE Proceedings, vol. 152, pp. 311-319, 2005.

[42] G. Holland, N. Vaidya and P. Bahl, "A rate-adaptive MAC protocol for multi-hop wireless networks," pp. 236-251, 2001.

[43] I. Haratcherev, J. Taal, K. Langendoen, R. Lagendijk and H. Sips, "Automatic IEEE 802.11 rate control for streaming applications," Wireless Communications and Mobile Computing, Special Issue on Radio Link and Transport Protocol Engineering for Future-Generation Wireless Mobile Data Networks, vol. 5, pp. 421–437, 2005.

[44] J. C. Bicket, “Bit-rate Selection in Wireless Networks”, M.S Thesis, MIT, Feb. 2005 http://madwifi.org

[45] B. Sadeghi, V. Kanodia, A. Sabharwal and E. Knightly, "Opportunistic media sccess for multirate ad hoc networks," in 2002, pp. 24-35.

[46] I. Haratcherev, K. Langendoen, I. Lagendijk and H. Sips, "Hybrid rate control for IEEE 802.11," in 2004, pp. 10–18.

[47] S. H. Y. Wong, S. Lu, H. Yang and V. Bharghavan, "Robust rate adaptation for 802.11 wireless networks," in 2006, pp. 146-157.

[48] E. Carlson, C. Prehofer, C. Bettstetter, H. Karl and A. Wolisz, "A Distributed End-to-End Reservation Protocol for IEEE 802.11-Based Wireless Mesh Networks," 11, ,Selected Areas in Communications, IEEE Journal on 24, 2018-2027, 2006.

[49] E. Carlson, H. Karl, A. Wolisz and C. Prehofer, "Distributed allocation of time slots for real-time traffic in a wireless multi-hop network," in 2004.

[50] A. Hamidian, "Providing QoS Guarantees in Ad Hoc Networks through EDCA with Resource Reservation," available online: http://www.telecom.lth.se/Personal/alexh/report.pdf

[51] Yongho Seok, Jaewoo Park and Yanghee Choi, "Multi-rate aware routing protocol for mobile ad hoc networks," ,Vehicular Technology Conference, 2003. VTC 2003-Spring. The 57th IEEE Semiannual 3, 1749-1752 vol.3, 2003.

[52] M. Bandai and T. Watanabe, "An On-Demand Routing Using Signal Strength for Multi-Rate Ad Hoc Networks," IEICE TRANSACTION COMMUNICATIONS E SERIES B, vol. 90, pp. 2504, 2007.

[53] D. S. J. D. Couto, D. Aguayo, J. Bicket and R. Morris, "A high-throughput path metric for multi-hop wireless routing," Wireless Networks, vol. 11, pp. 419-434, 2005.

[54] R. Draves, J. Padhye and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks," in 2004, pp. 114-128.

[55] S. Kim, S. Lee and S. Choi, "The impact of IEEE 802.11 MAC strategies on multi-hop wireless mesh networks," ,2nd IEEE Workshop on Wireless Mesh Networks, 2006. WiMesh 2006. 38-47, 2006.

[56] J. C. Park and S. K. Kasera, "Expected data rate: An accurate high-throughput path metric for multi-hop wireless routing," University of Utah, Tech.Rep., Apr, pp. 05-006, 2005.

[57] B. Awerbuch, D. Holmer and H. Rubens, "The Medium Time Metric: High Throughput Route Selection in Multi-rate Ad Hoc Wireless Networks," Mobile Networks and Applications, vol. 11, pp. 253-266, 2006.

[58] R. P. Guimaraes and L. Cerda, "Improving Reactive Routing on Wireless Multi-rate Ad-hoc Networks," available online: https://upcommons.upc.edu/e-prints/bitstream/2117/1173/1/mr-aodv.pdf

[59] Z. Fan, "High throughput reactive routing in multi-rate ad hoc networks," 25, Electronics Letters 40, 1591-1592, 2004.

[60] L. Iannone and S. Fdida, "Can multi-rate radios reduce end-to-end delay in mesh networks? a simulation case study," in 2005.

[61] M. Genetzakis and V. A. Siris, "A contention-aware routing metric for multi-rate multi-radio mesh networks," in 2008, pp. 242-250.

[62] OPNET Modeler, OPNET Technologies Incorporation. www.opnet.com

Chapter 6

Conclusion and Future Work

In this thesis we focused on the IEEE 802.11 MAC and identified several issues which required further research. Various design strategies were investigated in the existing research literature and discussed in detail. Based on the findings and the requirements, we presented our design, followed by the proof of concept. To broaden the scope, we discussed the impact of MAC specific enhancements on the overall communication system in a multi-rate, multi-hop mobile ad-hoc network scenario. To integrate the *standalone* (layer-specific) enhancements we presented a cross-layer frame-work and proposed a routing protocol which relies on joint parameters belonging to several layers. Various performance tests affirmed the advantages of using the cross-layer framework and integration of cross-layer information.

The key challenges addressed and the solutions proposed through the research presented in this thesis are as follow:

6.1 Design of a rate-adaptation protocol

6.1.1 Challenge-1

The IEEE 802.11 standard conformant wireless communication devices are capable of transmitting at various transmission rates by changing the underlying modulation schemes. Changing the transmission rates is required accordingly to the changes in the SNR. There are a many factors which cause variation in SNR. Therefore, adaptive changes of the transmission rate are extremely important for efficient communication. Although, rate-adaptation protocol is of significant importance, the IEEE 802.11 standard does not provide standard specification for a rate-adaptation technique. The standard, however, specified mandatory rules for devising a rate adaptation technique. As a result of the lack of standard specification, there are a number of techniques proposed by manufacturers of the Standard conformant devices and from independent researchers. Each of the proposed approaches addresses the issue of rate-adaptation from a different perspective, and has different implications on the communication.

The research challenge is to design a rate-adaptation protocol which is:

1. Highly responsive to variations in the channel quality (SNR).
2. Can detect the actual reason of frame losses.
3. Is able to determine the actual (closest possible) duration of channel variations; and thus avoid *blind* rate-up attempts.
4. The rate-adaptation protocol should not require any changes to the Standard frame formats.
5. Due to the time-constrained nature of task of rate-adaptation, such schemes should not be computationally expensive and processing and computational overhead should be minimized.

6.1.2 Contributed Rate-adaptation protocols (solution to challenge-1)

In this thesis, we present two rate-adaptation protocols. As discussed in detail in the thesis, there are two classes of rate-adaptation techniques: *the sender-side schemes* and the *receiver-side schemes*. We proposed CLRA, which is a sender side rate-adaptation technique and does not require any feedback from its receiver. The second contribution is a receiver-side (closed-loop) rate adaptation technique called MutFed. MutFed relies on the receiver feedback in performing rate-adaptation.

The key contributions and some of the aspects of the proposed design are as follow:

1. A frame-failure-statistics based rate adaptation solution which uses an *on-demand incremental* strategy for selecting a rate-selection threshold. This solution is based on a cross-layer communication framework, where the rate-adaptation module involves information to/from the Application layer along with relevant information from the MAC sub-layer while making a decision.

   a. The on-demand incremental strategy avoids the chances of retransmissions.
   b. The rate-control function works in close coordination with the Application layer requirements.
   c. The underlying MAC sub-layer's timing constraints and medium congestion is taken into account while performing rate-adaptation.
   d. A feedback mechanism is used to communicate the underlying layers' status and capability information to higher layers.
   e. A novel mechanism of frame-loss differentiation is introduced to distinguish between frames lost because of erroneous transmissions (lost frames) from frames which are not received because of simultaneous transmission by other stations.
2. An SNR-based rate-adaptation scheme, called MutFed, relying on mutual feedback between a transmitter and receiver pair.
   a. MutFed uses a novel mechanism of conveying feedback between communication peers, without the need for changing the Standard frame formats.
   b. The proposed rate-adaptation scheme is highly responsive to variations in communication-channel's quality.
   c. MutFed uses an extremely efficient and highly responsive frame-loss differentiation mechanism for distinguishing *corrupted frames* from *collided frames.*

CLRA estimates an instantaneous value of the higher transmission rate which a rate adaptation scheme should attain under the existing conditions. The motivation is partly based on the fact that CLRA should avoid the chances of encountering frame-losses (and thus retransmissions) to the best; and partly the motivation comes from the fact that even in ideal channel quality conditions, it is not a requirement to transmit at the highest possible transmission rate. Combining the two reasons, we came up with a solution to determine a runtime requirement of application's generated traffic, then determine, the underlying medium access constraints (and thus estimate the contention delays), and then select such a value which is most suitable and at the same time minimizing the chances of retransmissions. CLRA is self-adjusting according to the medium contention levels. Therefore, a station whose application generated traffic is very low would not be able to penalize the rest of the stations by operating at a lower rate. In this situation, the medium contention becomes higher and thus the CLRA limiting values are also incremented. CLRA uses a window RTS/CTS exchange for differentiating *frame collisions* from *frame corruption*. The mechanism is effective in determining the presence of hidden nodes and then adjusting the rate-adaptation scheme. However, the use of RTS/CTS is a communication overhead. The proposed frame-loss-differentiation mechanism uses the RTS/CTS procedure effectively to minimize the overhead.

MutFed is a receiver side-SNR based rate-adaptation scheme which uses receiver's feedback while selecting a transmission rate for future transmissions. Unlike the previous receiver-side (closed-loop) rate adaptation schemes MutFed do not require modification of the standard frames. MutFed is highly responsive when compared to statistics based rate-adaptation schemes and converge quickly to the most suitable transmission rate under prevailing channel quality. MutFed does not rely on frame deliver statistics and thus the rate-up and rate-down attempts are more realistic and guided by the actual SNR. MutFed also introduces a novel mechanism of delivering the feedback information from the receiver back to the transmitter. The rate-selection feedback, which is suggested by a receiver, is delivered without any modification to the standard frames as has been suggested in the previous research literature. MutFed uses a novel and extremely efficient mechanism for differentiation of frame losses. The frame loss differentiation is performed without the use of RTS/CTS exchange, thus avoiding communication overhead. Moreover, loss differentiation is achieved just with the help of three frames. The loss-differentiation mechanism enables the rate-adaptation process to work in an intelligent manner.

However, it is important to notice that MutFed is a closed-loop rate-adaptation scheme which requires a periodic input from a receiver, therefore, MutFed is only applicable in networks in which all stations are MutFed conformant.

6.1.3 Challenge-2

There are three key issues, which constitute this challenge:

1. Majority of MANET routing protocols use broadcasts during route discovery process. The route selection in most conventional routing protocols is usually based on the minimum hops between the source and destination. In a route with minimum hops between the source and destination, the distance between neighbouring nodes is usually the longest and may cause creation of *communication gray zones*, where stations which are marked as '*reachable neighbours'* using the broadcast frames would not be accessible during normal, unicast (possibly using higher transmission rate) communication. Some factors which contribute in the formation of *communication gray zones* are: multi-rate transmission capability of the physical layer (PHY), no acknowledgment policy in the IEEE 802.11 MAC when broadcast frames are sent, the small size of broadcast frames which increases the transmission success probability of broadcast frames over unicast data (possibly larger) frames, and the link fluctuations which can disrupt border line stations. When rate adaptation schemes are used at the MAC layer the routes discovered at the lowest rate (used for broadcast frames) may not be accessible. To overcome this problem and to better utilize the potential of the lower layers, a routing protocol should be aware of the rate-adaptation at the PHY.
2. In a MANET (based on IEEE 802.11 conformant stations), having higher station density a station may have to freeze its backoff counters a number of times before it finds an opportunity to transmit on the medium. Transmission errors and station density would account for higher medium occupancy by neighbouring stations and as a result every station would experience higher contention delay for accessing the medium. Likewise, stations using lower transmission rate can penalize neighbouring stations by causing higher medium contention. Therefore, a station may be able to transmit at a higher transmission rate; however, because of medium occupancy the medium access delay would be higher. In such situation, a route having higher throughput links would encounter higher end-to-end delay; the routing protocol should be capable to take into account this information while selecting a route in a MANET.

3. The IEEE 802.11e introduces four different MAC level queues for each of the four access categories. The standard specifies maintenance of service priority within the queues; which implies that frames from a higher priority queue would be serviced more frequently than frames belonging to lower priority queues. Such an enhancement at the MAC sub-layer introduces uneven queuing delays. Conventional/existing routing protocols are unaware of such queue prioritization and as a result these factors are not considered which result in severe performance deterioration for frames belonging to lower priority queues. It is important to consider hop-by-hop (or a link) MAC level information while constructing an end-to-end routing strategy for a multi hop wireless network.

6.1.4 MADV routing protocol (solution to challenge-2)

In this thesis we proposed a routing protocol called Medium Aware Distance Vector (MADV). MADV formulates a routing metric which combines variations in the transmission rates, contentions levels and queuing delay at every link. The research in connection with the challenges mentioned above investigated the significance of variation of the transmission rate, the use of a rate-adaptation scheme at the MAC layer and its effects on the performance of routing protocols and highlighted the key issues specific to routing protocols. Moreover, the significance of medium occupancy, PHY and MAC level factor contributing to this phenomenon and the effects of medium contention on the performance of route selection. An in depth analysis of the IEEE 802.11e inter-queue access priority, the resultant queue delay variations and their effects on routing strategy. These three factors lead to the formulation of the MADV routing metric. MADV is inherently based on the AODV operational characteristics. The key differences of MADV are as follow:

1. MADV uses a route selection metric based on the per-AC-per-destination in a MANET. This is completely different than the existing/conventional routing protocols and it essentially means that a source station can have different route entries based on the ACs for the same destination. The advantage of ACs-specific route entries is that MAC-level queue prioritization can be effectively taken into account for the whole route, and the route selection is performed based on the lowest metric value per AC from the available routes to a destination.

2. MADV uses a cross-layer information exchange frame work for communication of MAC and PHY specific parameters for inclusion in the routing protocol route metric.

3. MADV incorporates the underlying MAC and PHY specific parameters and the overall route selection is based on the actual state of the medium.

MADV decouples the conventional notion of hop-count from the underlying network performance. Parameters like the queuing and contention delay are dependent on the selection of appropriate *stations* along the route which offer minimum queuing and contention delays. MADV is an alternative approach of using the underlying MAC-layer parameters in construction of the routing metric and our objective with the use of MADV is not to implement yet another routing protocol for routing in mobile ad-hoc multi-hop networks, but to integrate the MAC layer enhancements into the routing protocols in the form of the MADV metric for performance enhancements. The MADV uses the same approach as AODV for route discovery and maintenance, with slight modification for integrating the new route metric for route selection. Route discovery is performed with the help of route request (RREQ) and route reply (RREP) initiated by the source and destination/intermediate stations respectively. We believe that by integrating the MAC layer parameters, which are significantly important for determining the quality of communication in terms of various QoS parameters (e.g. delay, throughput, frame error rate etc), we can significantly improve the performance of existing routing protocols. The MADV-metric can be incorporated into various routing protocols and its applicability is determined by the possibility of provision of MAC dependent Q.C.P parameters which are used to determine the hop by hop MADV-metric values.

6.2 Future research

The research presented in this thesis will be extended in future, in several ways.

1. The ongoing and future research is primarily focused on broadening the impact of the rate adaptation schemes' designs (presented in the thesis) by combining such schemes to coexist with Adaptive Transmit Power Control (ATPC) at the MAC layer.

There is a tremendous research focus on efficient use of the limited battery resources in wireless communication stations. Adaptive adjustment of transmission power has the potential to reduce the overall power consumption in wireless stations. Ideally, a transmitter should be configured to use the highest transmission rate and the lowest possible transmission power. Using lower transmission power also reduces interference and network capacity can be increased. However, in DCF mode of operation, if stations are allowed to transmit at variable power, the number of hidden nodes are more likely to increase, and retransmission (as a result of collisions) are highly likely to occur, which in turn will consume more power [1, 2]. Likewise, high rate transmissions require a higher SNR at the receiver for successful reception. Therefore, increasing the transmission rate and reducing the transmission power can result in frame losses.

To address the issue of adaptive-TPC, a number of mechanisms has been proposed [1- 9]. In order to cope with the hidden node problem with the use of adaptive-TPC, [2] suggests the use of RTS/CTS frames before the exchange of every data frame. The CTS frames are sent at a highest power (and a lower transmission rate as mandated by the IEEE 802.11 standard for sending control frames), thus increasing the transmission range of the CTS frame. The actual data frames are then sent at the appropriate transmission power based on the channel state information (CSI). Similar schemes are proposed in [5, 6, 7], where the authors in [7] proposed the adaptive tuning of transmit power and frame size according to the CSI. For PCF-based BSS, [4] suggests a mechanism where the Point Coordinator (PC) sends rate and power information in the SERVICE field of a MAC frame to a station. The station after receiving the frame can estimate the path loss between itself and the PC and using this estimation, select the most suitable transmission rate and power for data frame transmission. The same authors in [8] suggested the use of computing offline values for rate-power and then selecting the appropriate values as simple table lookup at run time. The authors in [9] proposed two different strategies for power and rate adaptation, based on the BSS scenario. According to [8], a station can either operate in a *high performance mode* where the transmission is done at the highest possible rate and the highest power, or it can use the *lower power mode* to focus on energy saving at the expense of compromising performance in terms of throughput and delay. The existing mechanisms of adaptive-TPC rely on either the transmitter side CSI estimation mechanisms (using counters frame transmission status) or using receivers' feedback. Therefore, such schemes have similar performance and implementation issues as the rate-adaptation schemes which rely on similar mechanisms of CSI estimation (as mentioned in chapter-3) .

The rate-adaptation schemes presented in this thesis use fixed transmission power. Lowering the transmission rate in response to channel quality variation may not necessitate any change in the transmission power and vice versa. However, these parameters are inter-related and significantly affect the performance of MAC layer. The coexistence of rate-adaptation schemes and adaptive-TPC at the MAC is a key research challenge, which will be addressed in the future work.

2. In future, the proposed rate-adaptation schemes will be tested on a Multi Band Atheros Driver for WiFi (MADWIFI) based WLAN testbed.

3. The proposed rate-adaptation schemes along with the adaptive-TPC mechanisms will be extended to the IEEE standards for Wireless Personal Area Networks (WPANs, IEEE 802.15).

6.3 References

[1] S. D. Gray and V. Vadde, Throughput and Loss Packet Performance of DCF With Variable Transmit Power , IEEE 802.11-01/227, May 2001.

[2] Daji Qiao, S. Choi, A. Jain, K.G. Shin, "Adaptive transmit power control in IEEE 802.11a wireless LANs," Vehicular Technology Conference, 2003. VTC 2003-Spring. The 57th IEEE Semiannual , vol.1, no., pp. 433-437 vol.1, 22-25 April 2003.

[3] J. Jelitto, J, A.N. Barreto, L.T. Hong, "Power and rate adaptation in IEEE 802.11a wireless LANs," Vehicular Technology Conference, 2003. VTC 2003-Spring. The 57th IEEE Semiannual , vol.1, no., pp. 413-417 vol.1, 22-25 April 2003.

[4] D. Qiao, S. Choi, A. Soomro, and K. G. Shin, “Energy-efficient PCF operation of IEEE 802.11a wireless LAN,” in Proc. IEEE Conf. Computer Communications, pp. 580-589, June 2002.

[5] J. Gomez, A.T. Campbell, M. Naghshineh, C. Bisdikian, "Conserving transmission power in wireless ad hoc networks," Ninth International Conference on Network Protocols, 2001, vol., no., pp. 24-34, 11-14 Nov. 2001.

[6] S. Agarwal, R.H. Katz, S.V. Krishnamurthy, S. K. Dao, "Distributed power control in ad-hoc wireless networks," 12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, vol.2, no., pp.F-59-F-66 vol.2, Sep/Oct 2001.

[7] J. P. Ebert, A. Wolisz, "Combined tuning of RF power and medium access control for WLANs," IEEE International Workshop on Mobile Multimedia Communications. (MoMuC '99) 1999 , vol., no., pp.74-82, 1999.

[8] D. Qiao, S. Choi, A. Jain and K. Shin, MiSer: An optimal low-energy transmission strategy for IEEE 802.11a/h, in: Proceedings ACM MobiCom pp. 161–175, 2003.

[9] P. Chevillat, J. Jelitto, and H. L. Truong. Dynamic Data Rate and Transmit Power Adjustment in IEEE 802.11 Wireless LANs. Intl. Journal of Wireless Information Networks, 2005.

List of research papers

Accepted papers:

1. S. Khan, S. A. Mahmud, K. K. Loo and H. S. Al-Raweshidy, "A cross layer rate adaptation solution for IEEE 802.11 networks," *Computer Communications,* vol. 31, pp. 1638-1652, 5/25. 2008.
2. S. A. Mahmud, S. Khan and H. Al-Raweshidy, "Meshed high data rate personal area networks," IEEE Communications Surveys & Tutorials, vol. 10, pp. 58-69, 2008.
3. S. Khan, S. A Mahmud and H.S Al-Raweshidy, "Rate adaptation for multi-rate wireless stations: A cognitive approach," in ICT Mobile Summit, 2008.
4. S. A. Mahmud, S. Khan and H. Al-Raweshidy, "Capacity Issues in Meshed High Data Rate WPANs," *Advanced Information Networking and Applications - Workshops, 2008. AINAW 2008. 22nd International Conference on,* pp. 1285-1290, 2008.
5. S. Khan, S. A Mahmud and H.S Al-Raweshidy, "A rate-adaptive MAC for IEEE 802.11 networks," in *Sixth Annual Conference on Communication Networks and Services Research,* 2008.
6. S. A. Mahmud, S. Khan and H. Al-Raweshidy, "Meshed high data rate personal area networks," Communications Surveys & Tutorials, IEEE, vol. 10, pp. 58-69, 2008.
7. S. Khan, S. A Mahmud and H.S Al-Raweshidy, "A Cross-Layer Solution for Dynamic Selection of Data-Rate in WLAN using Traffic Differentiation," *Mobile and Wireless Communications Summit, 2007. 16th IST,* pp. 1-5, 2007.
8. A. N. Al-Khwildi, S. Khan and H. Al-Raweshidy "On-Demand Link Weight Routing Protocol with Cross-Layer Communication," *Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on,* pp. 1-5, 2007.
9. S. Khan, S. A Mahmud, S. Khan, W. Yeo and F. Cecelja, "Generalized architecture for converged heterogeneous networks," in *The 16th WWRF Meeting,* 2006.

10. S. A. Mahmud, S. Khan, S. Khan and H. Al-Raweshidy, "A comparison of MANETs and WMNs: commercial feasibility of community wireless networks and MANETs," *Proceedings of the 1st International Conference on Access Networks,* 2006.

11. S. Khan, S. A Mahmud and H.S Al-Raweshidy, "Supplementary Interworking Architecture for Hybrid Data Networks (UMTS-WiMAX)," Computing in the Global Information Technology, 2006. ICCGI '06. International Multi-Conference on, pp. 57-57, 2006.

12. S. Jan, S. Khan, M. Li and H.S. Al-Raweshidy, "Distributed Execution of an Inherently Sequential Network Simulator", IEEE International Conference on Communication Software and Networks (ICCSN), 2009.

Papers currently under-review/ in press:

1. S. Khan, S. A. Mahmud, H. S. Al-Raweshidy, "Rate-adaptation for multi-rate IEEE 802.11 WLANs using mutual feedback between transmitter and receiver", IEEE Trans. Mobile Computing, 2009.

2. S. Khan, S. A. Mahmud, H. S. Al-Raweshidy, "A survey of rate-adaptation schemes for multi-rate WLANs", IEEE Communications Surveys & Tutorials, 2009.

3. S. Khan, S. A. Mahmud, H. S. Al-Raweshidy, "Medium-Aware Distance Vector (MADV) routing protocol for multi-rate, IEEE 802.11e based MANETs", *Computer Communications,* 2009.

4. S. A. Mahmud, S. Khan, H. S. Al-Raweshidy, "A Resource Allocation Strategy for Meshed High Data Rate WPANs", IEEE Comm. Letters. 2009.

5. H. F. AbuTarboush, S. Khan, R. Nilavalan, D. Budimir, H. S. Al-Raweshidy, "Reconfigurable Planar Inverted-F Antenna for Wideband Cognitive Radio", Microwave and Optical Technology Letters, 2009.