OPPORTUNISTIC SCHEDULING FOR NEXT GENERATION WIRELESS LANS

by

ERTUĞRUL NECDET ÇİFTÇİOĞLU

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of
the requirements for the degree of
Master of Science

Sabancı University
August 2006

OPPORTUNISTIC SCHEDULING FOR NEXT GENERATION WIRELESS LANS

APPROVED BY:

Assist. Prof. Dr. Özgür Gürbüz ....
(Thesis Advisor)

Assist. Prof. Dr. Özgür Erçetin ....

Assist. Prof. Dr. Mehmet Keskinöz ....

Assist. Prof. Dr. Albert Levi ....

Assist. Prof. Dr. Oğuz Sunay ....

DATE OF APPROVAL:

OPPORTUNISTIC SCHEDULING FOR NEXT GENERATION WIRELESS LANS

Ertuğrul Necdet Çiftçioğlu

EECS, MS Thesis, 2006

Thesis Supervisor: Assist. Prof. Dr. Özgür Gürbüz

Keywords: Scheduling, Wireless networks, Wireless communications

ABSTRACT

The multiuser diversity phenomenon has been exploited via opportunistic scheduling for increasing system throughput in wireless networks recently. Frame aggregation which increases the MAC efficiency is another method in enhancing system throughput. Opportunistic scheduling has been employed jointly with frame aggregation in order to maximize the system throughput in this work. The use of existing opportunistic schemes has been shown not to be optimal when frame aggregation is used, as applied in the IEEE 802.11n Wireless Local Area Network standard. New scheduling approaches which combine channel states with queue states have been proposed with the aim of maximizing the total network throughput. In addition to schedulers which select transmitted users according to instantaneous scheduling metrics, schedulers which maximize throughput over larger time intervals have been proposed. These schedulers utilize results obtained from the queuing model developed for 802.11n throughout this thesis. The proposed new algorithms are shown to offer significant improvement in network throughput over non-opportunistic and greedy schedulers through detailed simulations. The developed algorithms also provide a good compromise between throughput and fairness. The effect of incorporating relaying in schedulers applied for frame aggregation systems has also been analyzed.

YENİ NESİL KABLOSUZ YEREL AĞ BAĞLANTILARI İÇİN FIRSATÇI ÇİZELGELEME

Ertuğrul Necdet Çiftçioğlu

EECS, Master Tezi, 2006

Tez Danışmanı: Yard. Doç Dr. Özgür Gürbüz

Anahtar Kelimeler: Çizelgeleme, Kablosuz ağlar, Kablosuz iletişim

ÖZET

Kablosuz ağlarda son yıllarda üretilen iş miktarını arttırmak için fırsatçı çizelgeleme metodları çoklu kullanıcı çeşitliliği olgusundan faydalanmışlardır. Sistemde üretilen iş miktarını yükseltmek için başka bir yöntem de Ortam erişim kontolü(MAC) verimini arttıran çerçeve birleştime metodlarıdır. Bu çalışmada fırsatçı çizelgeleme metodları çerçeve birleştirme metodlarıyla birlikte uygulanmıştır. IEEE 802.11n IEEE 802.11n Kablosuz Yerel Ağlarında uygulanda da kullanılan çerçeve birleştirme ile birlikte var olan fırsatçı çizelgeleme metodlarının kullanımın optimum olmadığı gösterilmiştir. Ağdaki toplam üretilen iş miktarını en çoklamak için kanal durumlarını ve kuyruk durumlarını birlikte değerlendiren yeni çizelgeleme metodları geliştirilmiştir. Anlık çizelgeleme metriklerine göre iletim yapılacak kullanıcıyı seçen çizelgeleyicilerin yanı sıra daha uzun zaman aralıklarına göre üretilen iş miktarını en çoklayan çizelgeleme yöntemleri önerilmiştir. Bu çizelgeleyiciler 802.11n için bu tez boyunca geliştirilen kuyruklama modelinden çıkarılan sonuçlardan faydalanmaktadırlar. Önerilen yeni algoritmalar, ayrıntılı benzetim sonuçları sonucunda fırsatçı olmayan ve aç gözlü çizelgeleyicilere göre ağdaki toplam üretilen iş miktarında önemli artış göstermişlerdir. Geliştirilen algoritmalar ayrıca üretilen iş miktarı ve denkserlilik konusunda iyi bir seçim sunmaktadırlar. Çerçeve birleştiricili çizelgeleyicilerde röleleme kullanımının etkileri de incelenmiştir.

To my lovely family

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor Assist. Prof. Dr. Özgür Gürbüz for her invaluable guidance, encouragement, great patience and support throughout my thesis studies. I consider myself extremely fortunate to have the opportunity to work with an advisor of such understanding and excellent personality. I am also grateful to members of my thesis jury, Assist. Prof. Dr. Özgür Erçetin, Assist. Prof. Dr. Mehmet Keskinöz , Assist. Prof. Dr. Albert Levi and Assist. Prof. Dr. Oğuz Sunay for spending their valuable time and energy.

I would like to express my gratitude to Cisco Sytems University Research Program and The Scientific & Reseach Council of Turkey (TÜBİTAK) for funding this research and supporting me financially throughout my thesis studies.

Especially, I would like to thank my dear family for their unconditional and endless love and support. It was impossible to achieve anything without them. Last but not least, I would like to thank my aunt, who lives in Istanbul, for her support and all members of the COMNET laboratory, especially Engin Maşazade and Hülya Seferoğlu for creating a stimulating environment throughout lengthy working hours.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACK:	Acknowledgement
ADOS:	Aggregate Discrete Opportunistic Scheduling
AOS:	Aggregate Opportunistic Scheduling
AP:	Access Point
AS:	Angular Spread
BLACK:	Block Acknowledgement
BLAR:	Block Acknowledgement Request
BPL:	Bit Power Loading
BSS:	basic Service Set
CFP:	Contention Free Period
CQS:	Capacity Queue Scheduling
CRC:	Cyclic Redundancy Check
CSMA/CA:	Carrier Sense Multiple Access/ Collision Avoidance
CSMA/CD:	Carrier Sense Multiple Access/ Collision Detection
CTS:	Clear To Send
CW:	Collision Window
dB:	Decibel
DCF:	Distribution Coordination Function
DIFS:	Districuted Interframe Space
DSSS:	Direct Sequence Spread Spectrum
ESS:	Extended Service Set
FCS:	Frame Check Sequence
FHSS:	Frequency Hopping Spread Spectrum
HT:	High Throughput
GI:	Guard Interval
IAC:	Initiator Aggregation Control
IEEE:	Institute of Electrical and Electronics Engineers
IFFT:	Inverse Fast Fourier Transform
LQ:	Longest Queue
LTF:	Long Training Field
MAC:	Medium Access Control
MCS:	Modulation Coding Scheme

MFB:	Modulation Feedback
MIMO:	Multiple Input Multiple Output
MPDU:	Medium Access Control Layer Protocol Data Unit
MRA:	Multiple Receiver Aggregation
MRAD:	Multiple Receiver Aggregate Descriptor
MRMRA:	Multiple Response Multiple Receiver Aggregation
MRQ:	MCS Request
MRS:	Maximum Rate Scheduling
NAV:	Network Allocation Vector
NLOS:	Non Line-of-sight
OAR:	Opportunistic Autorate
OFDM:	Orthogonal Frequency Division Multiplexing
OPNET:	Optimized Network Engineering Tool
OSI:	Open Systems Interconnect
P-AG:	Predictive Scheduling with Access Guarantees
P-AOS:	Proportional Aggregate Opportunistic Scheduling
P-WF:	Predictive Scheduling with Time-domain Waterfilling
PAS:	Power Angular Spread
PHY:	Physical Layer
PCF:	Point Coordination Function
PIFS:	Point Coordination Interframe Space
PFQ:	Proportional Fair Queuing
PLCP:	Physical Layer Convergence Procedure
PMD:	Physical Medium Dependent
PPDU:	Physical Layer Data Unit
PSDU:	Physical Layer Service Data Unit
PSK:	Phase Shift Keying
RAC:	Responder Aggregation Control
RDG:	Reverse Direction Grant
RDL:	Reverse Direction Limit
RDR:	Reverse Direction Request
RDTID:	Reverse Direction Traffic Identifier
RIAC:	Relayed Initiator Aggregation Control
RPO:	Reverse Period Offset

RTS:	Request To Send
QAM:	Quadrature Amplitude Modulation
QoS:	Quality of Service
SIFS:	Short Interframe Space
SISO:	Single Input Single Output
SNR:	Signal-to-Noise Ratio
SRA:	Single Receiver Aggregation
SRPT:	Shortest Remaining Processing Time First
STF:	Short Training Field
SVD:	Singular Value Decomposition
TGnSync:	Task Group N Sync
TRQ:	Training Request
TXOP:	Transmission Opportunity
WLAN:	Wireless Local Area Network

CHAPTER 1

INTRODUCTION

Wireless access has been increasingly popular recently due to portability and low cost. Extensive research is being carried out to increase provided data rates of wireless networks to values comparable with wired networks. The emerging technologies for wireless local area networks (WLANs) are defined by the IEEE 802.11 standards, which started by 802.11b with Physical Layer (PHY) data rates of up to 11 Mbps and were enhanced in 802.11a/g to provide up to 54 Mbps with the introduction of Orthogonal Frequency Division Multiplexing (OFDM). In the newest WLAN standard group 802.11n, PHY data rates exceeding 200 Mbps are provisioned with the realization of multiple input multiple output (MIMO) techniques [1]. MIMO systems are based on the presence of multiple antennas at the transmitter and receiver ends of the communication link, which make use of spatial diversity and yield significant increase in system capacity. On the other hand, the actual throughput experienced by the WLAN users is considerably lower than the PHY data rates. Since the transmission media is shared, efficient multi/Medium Access Control (MAC) is the key to provide desired high data rate services.

In the new MAC standard 802.11e and high data rate draft standard 802.11n, MAC efficiency is to be enhanced via the frame aggregation concept [2], [3]. In WLANs there are mainly two types of packets that are transmitted: control packets for coordination among stations and data packets carrying desired data payload. Data packets are transmitted at a data rate depending on the channel quality of the served user, while control packets are transmitted at the lowest data rate, so as to be decoded by all users. Even though the control packets are much smaller than the data packets, the time wasted due to control packet transmission is not negligible [2]. In addition to control packet transmission, another source of overhead is the Physical Layer Convergence Procedure (PLCP) overhead that is added to all packets. In order to reduce the relative percentage of the time loss due to packet overhead and MAC coordination, the method of frame aggregation has been introduced, where multiple MAC layer protocol data units (MPDUs) are transmitted in one physical layer protocol data unit (PPDU). Frame aggregation is shown to offer significant improvement in MAC efficiency especially when the packet sizes are small [2] .

In a multiuser communication system with an infrastructure network topology, scheduling is the mechanism that determines which user should transmit/receive data in a given time interval. Scheduling is an essential element of the MAC layer due to its effect on the overall behaviour of the system. Opportunistic scheduling algorithms [4], [5], [6], [7] maximize system throughput by making use of the channel variations and multiuser diversity. The main idea is to favour users that are experiencing the most desirable channel conditions at each scheduling instant [4]. While enhancing the throughput, an algorithm based only on transmitting over users with the best channels may cause some users to experience unacceptable delays. Unfairness of greedy algorithms can be resolved via the proportional fair approach, where throughput and fairness performance are compromised by favouring users with relatively good channel conditions with respect to their own history [7].

In this thesis, we combine aggregation and opportunistic scheduling approaches to further enhance the throughput of next generation WLANs. Specifically, the system proposed in the draft standard for 802.11n by TGnSync group [1] was considered. We argue that aggregation can dramatically change the scheduling scenario: A user with a good channel and a long queue may offer a higher throughput than a user with better channel conditions but shorter queue. Hence, the statement that always selecting the user with the best channel maximizes throughput is not valid anymore. In this work, first new queue aware scheduling schemes that take into account the instantaneous channel capacities and queue sizes simultaneously are proposed. Queue aware schedulers such as Aggregate Opportunistic Scheduling (AOS), Aggregate Discrete Opportunistic Scheduling (ADOS), Proporional Aggregate Opportunistic Scheduling (P-AOS) and Capacity Queue Scheduling (CQS) are proposed. Through detailed simulations, we evaluate the performance of our algorithms in comparison with known algorithms from literature [4], [7], [8], [9]. We also consider non-opportunistic schemes [1] as well as MAC with no aggregation. Our results indicate that proposed algorithms offer significant gains in throughput while permitting relatively fair access. We also improve our algorithms AOS and CQS with the principle of relaying transmission.

Later on, we propose schedulers which do not only perform scheduling depending on instantaneous channel and queue states, but aim to maximize throughput over a larger time scale. In order to design such schedulers, statistical evolution of the queue states is required. Hence, we model the 802.11n MAC using queuing theory by extending the bulk service model [10]. Utilizing the outcomes of the queuing model, two schedulers, Predictive Scheduling with Access Guarantees (P-AG) which provides QoS guarantees and another Predictive Scheduling with Time-domain Waterfilling (P-WF) which is the based on the application of the waterfilling principle are developed. These schedulers, which we denote as "*controlled-access*" schedulers, further improve performance with respect to queue aware schedulers.

The rest of the thesis is organized as follows: In Chapter 2, a general overview of the MAC and PHY layers of the 802.11 and 802.11n standards is provided, introducing concepts related with frame aggregation, opportunistic scheduling and MIMO systems as well as general protocol standards. Our proposed schedulers which depend on instantaneous channel and queue states are demonstrated in detail in Chapter 3, where extensions of the schedulers to relaying are also provided. The modeling of the 802.11n MAC using queuing theory and statistical based long term schedulers which control access proportions of users throughout the scheduling duration are presented in Chapter 4. In Chapter 5, we present the simulation model including network, air interface and channel model, capacity calculation and MAC framework. In Chapter 6 we present our results and performance evaluation, where we compare our proposed scheduling algorithms with existing schedulers from the literature. Finally, Chapter 7 involves our conclusions and directions for future work.

CHAPTER 2

BACKGROUND

2.1 IEEE 802.11 Wireless LANs

Wireless networks have attracted great interest due to advantages such as mobility low deployment costs. Even though the initial idea of Wireless LANs emerged in late 1970s [11], the first standard which secured the market was completed in 1997. Wireless LAN standards have been developed by the IEEE 802.11 comittee. In this section, the basics of the initial 802.11 standards[12] will be discussed.

Despite the advantages offered by deploying Wireless LANs, the development of the underlying technology for Wireless LANs are much more difficult due to the characteristics of the wireless medium. Furthermore, additional constraints such as limited power arise when user mobility is considered. The most important disadvantage of the wireless medium is the presence of path loss, fading and shadowing, which cause the transmitted signal to decay while propagation before reception.

2.1.1 IEEE 802.11 Architecture

Similar to the cellular phone network infrastucture, the 802.11 network is divided into cells in infrastructure mode. The cells are called a Basic Service Set (BSS), and the traffic is directed through access points (AP). In addition to the infrastructure mode where the access point has the capability to control network operation, the ad-hoc mode is also allowed with the absense of an access point . The access points are connected with a backbone distribution system, mostly wired, to form groups of BSS called the extended service set (ESS). ESS can be connected to a wired network through a device called a portal.

2.1.1.1 MAC Sublayer

In ethernet, which is the most common wired network standard, the deployed MAC protocol for medium access is the Carrier Sence Multiple Access/ Collision Detection (CSMA-CD) protocol. In CSMA/CD, a user which has data to transmit senses the medium. If the medium is idle, the station transmits data. Otherwise, the station defers transmission to a later time depending on the exact version of CSMA/CD. Even though carrier sensing reduces the likelihood of collisions due to simultaneous transmissions, collisions may still occur. If collision detected, the transmissing stations abort their transmission to avoid further time and power loss. Colliding stations defer transmission and sense the medim again after a time depending on the exponential backoff algorithm[13].

Even though CSMA/CD is satisfactory for wired LANs, it is not applicable to wireless LANs due to two main reasons. First, it is not possible to transmit and receive for carrier sense simultaneously at the same frequency band, leading to the requirement of full duplex radio, which is expensive. Second, the carrier sense mechanism may not be adequate for the wireless medium since a station may not detect presence of ongoing transmission due to limited coverage caused by path loss.

The medium access method used in Wireless LANs is the Distributed Coordination Function (DCF) which is based on the Carrier Sense Multiple Access / Collision Avoidance (CSMA/CA) protocol. In the CSMA/CA protocol, a station which has data to transmit senses the medium. If the medium is busy, it defers transmission and selects a Contention Window(CW) , which is a duration which must elapse until the station is allowed to transmit. The duration of the CW is the duration of a slot time multiplied by randomly selected value selected from the CW set . The CW is freezed until medium is sensed idle and a duration called Distributed Inter Frame Space (DIFS) passes. If the mediıum is still idle after DIFS, the contention window is decremented until new transmission activity is detected in the medium. When the contention window of a user vanishes to zero, the station transmits data. If the receiving station responds with an Acknowledgement (ACK) packet, the packet was delivered successfully. However, if an ACK is not received within the expected duration, the transmitter assumes the packet was not transmitted successfully and decides to retransmit. For retransmission, the station should select a contention window duration according to the exponential backoff algoritm. The exponential backoff algorithm causes the contention window to be selected from a set with higher upper values after collisions. After each collision, the maximum size of the seed of CW in terms of slot time is doubled until it reaches an upper limit of 1024. Stations which have to defer due to sensing the medium busy freeze their contention windows until the medium is sensed free again and an Inter Frame Space duration has elapsed. The existing contention windows are decremented when the medium is idle afterwards.

In 802.11, the hidden terminal problem is defined as the problem in which stations may not hear each other so stations sense the medium idle and they transmit, resulting in collision at the intersecting regions. An extended version of CSMA/CA was designed to overcome the hidden terminal problem which uses an initial handshake with the receiving station . After sensing the medium idle for DIFS, the transmitting station first sends a packet called Request-to-Send (RTS) to gain permission from the receiver . If the receiver also senses the medium idle, after waiting for a Short Inter Frame Space (SIFS) it responds with a Clear-to-Send(CTS) packet which indicates that data can be transmitted. All other stations receiving the CTS packet defer transmission. After receiving the CTS packet, data is transmitted after waiting for another SIFS duration. Finally, if the reception is successful, the ACK is sent to the transmitting station after waiting for a SIFS duration. Another extension is to introduce a virtual carrier sensing mechanism in addition to the physical carrier sensing. Virtual carrier sensing utilizes duration information in packets by setting a timer called Network Allocation Vector (NAV), which indicates the time that must elapse until the medium will be available for sampling to check idle status . The remaining stations defer transmission until both the physical carrier sensing indicates the channel is not busy and the NAV duration is finished. The resulting protocol operation is shown in Figure 2.1.

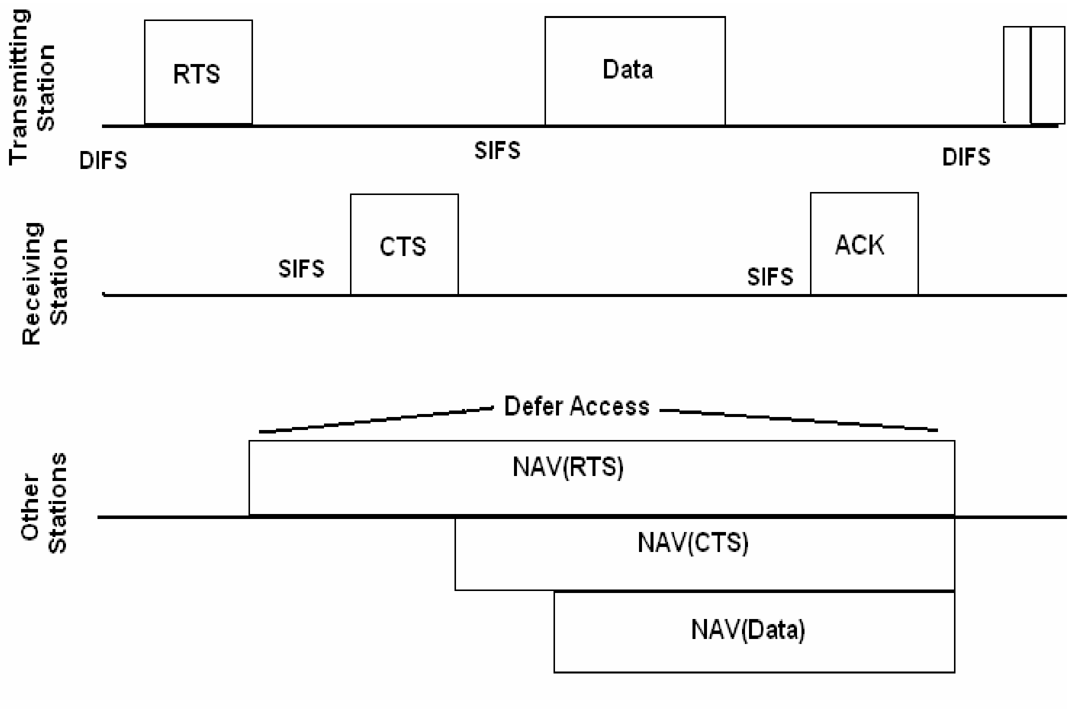


Figure 2.1 Timing diagram of 802.11 access mechanism.

In addition to the Distribution Coordination, an optional contention-free access method called Point Coordination Funciton (PCF) has been defined for services with high QoS requirements, such as voice and video. PCF is used in coexistence with DCF and the frequency of using PCF is determined by the contention free period (CFP) repetition interval. During the CFP, the access point polls stations with time-bounded data. The access point gains control for polling by accessing the medium after waiting for a shorther inter frame space called Point Coordination IFS (PIFS).

There are three types of packets in 802.11. Data frames carry the payload data which is to be transmitted. Control frames are used for medium access coordination such as RTS, CTS and ACK frames. Management frames are concerned with issues such as AP association and dissassociation, timing and synchronization , authentication and deauthentication.

2.1.1.2 PHY Sublayer

In 802.11, the physical layer tasks are partitioned into two sublayers. The Physical Layer Convergence Procedure (PLCP) sublayer converts the packet coming from the MAC layer to a format suitable for transmission. On the other hand, the Physical Medium Dependent (PMD) sublayer is responsible with transmission over the wireless channel. The initial standard provided three options for the physical layer, Frequency Hopping Spread Spectrum (FHSS), Direct Sequence Spread Spectrum (DSSS) and the infrared. The frequency band allowed for operation is 2.4 GHz- 2.4835 GHz. The 802.11b standard offered data rates up to 11 Mbps. In 802.11a, Orthogonal Frequency Division Multiplexing (OFDM) is deployed with data rates offered upto 54 Mbps.

2.2 Next Generation WLANs – IEEE 802.11n

The IEEE 802.11n is develped by the Task Group N (TGn). While previous task groups have aimed to increased peak throughput , or data rate, TGn intends to achieve at least 100 Mbps net throughput in the MAC layer after considering the effects of overhead, preambles and interframe spaces. The TGnSync proposal[1] has proposed the main standard specifications. In order to increase net throughput, enhancements in both physical and MAC layers have been proposed. MAC efficiency has been improved with the frame aggregation concept[2] [3], and physical layer data rates are mainly improved by deploying Multiple Input Multiple Output (MIMO) transmission and using channel bandwiths of 40 MHz in addition to the 20 MHz channels used in previous standards. In this section these enhancements are presented in more detail.

2.2.1 MAC Layer Enhancements

2.2.1.1 Frame Aggregation

Single Receiver Aggregation(SRA)

In previous 802.11 standards, data transmission is preceded by an RTS-CTS exchange. An ACK is also required. Typically, even though the length of control packets are much smaller than data payload in terms of bits, the transmission of control packets is carried out in a lowest common basic rate so that each station can decode the packets and adjust their behaviours accordingly. On the other hand, data packets are sent at a higher adaptive data rate depending on channel conditions. Since transmission duration is inversely proportional with data rate, the even if control packets are much smaller than data packets their transmission delays comparable with packet transmission times, so throughput is severely effected [2][14]. Note that there are also inter frame spaces and physical layer overhead such as the PLCP header. The overall MAC efficiency, which we define as the ratio of actual throughput to channel data rate increases with increased data packet length [2]. Efficiency is lower when high data rates are used since the data payload is transmitted much faster, while control overhead is fixed.

In order to improve MAC efficiency, the concept of frame aggregation was introduced where multiple MPDUs are transmitted in one physical frame. By doing so, the proportion of time wasted to overhead over all the transmission duration is reduced since the data duration and the amount of data sent is increased. Frame aggregation has emerged as a very crucial element in order to achieve high MAC-layer throughput, which is the main goal of IEEE 802.11n.

We consider a time division system where only one user is served at a given time period, limited by a duration called transmission opportunity (TXOP). Each user is allocated a separate queue at the AP. As defined by 802.11n draft standard, within a TXOP, a two-way handshake with frame aggregation can be performed as shown in Figure 2.2 [1]. Considering the downlink operation, when the AP has to send data to a user, it sends a Request-to-send (RTS)-like packet called Initiator Aggregation Control (IAC) is responded by a Clear-to-send (CTS)-like packet called Responder Aggregation Control (RAC). The size of the frame aggregation depends both on the queue size, the maximum limit on aggregate size, and the maximum allowed number of MPDUs that can be transmitted within the TXOP limit. A control packet called Block ACK Request (BLAR) is transmitted in the same frame with the aggregated data. As its name implies, the purpose of the BLAR packet is to request information from the receiving station about the reception status of the packets forming the aggregation. At the end of packet transmission, the responding station replies with a modified ACK packet, called Block ACK (BLACK) containing information about the reception status of packets in aggregation, which depends on the channel conditions associated with that user. If any of the packets in the aggregation are not correctly received, the undelivered packets are retransmitted as an aggregation.

The data packets are transmitted at a selected adaptive transmission rate during the transmission sequence. The selected data rate is preserved throughout the transmission of data MPDUs in the transmission sequence, and is selected before aggregation transmission according to channel conditions via rate adaptation, which will be discussed in Chapter 5. The control packets, including the BLACK and Block ACK Request (BLAR) packets, are transmitted at the basic rate, so that all stations can decode these packets, and defer until the end of the transmission. When all of the packets in the aggregation are delivered successfully, a new transmission sequence is selected.

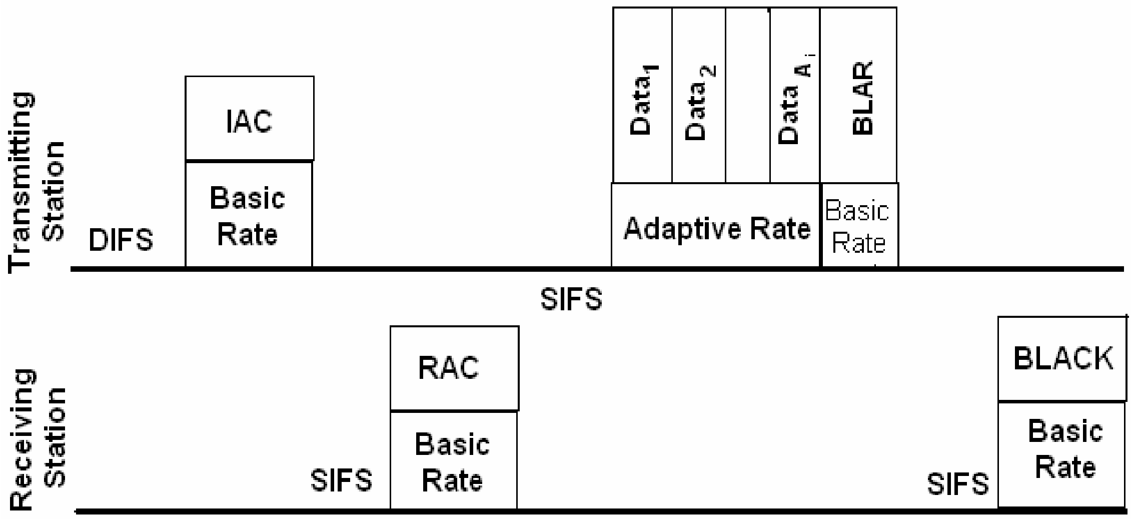


Figure 2.2: Example of frame aggregation exchange transmission.

At each transmission sequence, the AP transmits to a selected station using frame aggregation as explained. Station selection is to be done according to a scheduling algorithm. Details of widely used schedulers are presented in Section 2.3 , but briefly the default scheduler proposed for 802.11n compares the sizes of the queues of data to be transmitted corresponding to each station in terms of packets and selects the user with the largest queue size[1].

In addition to the aggregation transmission sequence explained, extensions of 802.11n supporting bi-directional data transfer have been proposed [3]. In bi-directional transmission, an initiator station may grant a Reverse direction grant (RDG), which is duration allowed for transmission from the receiver station to the initiator. If the receiver has any packets to send to the initiator station, it calculates if any packets can be sent together with a BLACK and RAC within the RDG duration.

Multiple Receiver Aggregation (MRA)

In contrast to single destination aggregation where the MPDUs destined to one particular station are combined in one PPDU, here multiple destination aggregation systems/ multiple receiver aggregation (MRA) are formed by groups of MPDUs destined to multiple receivers [1]. The frame exchange diagram for a MRA for two receivers can be shown as follows:

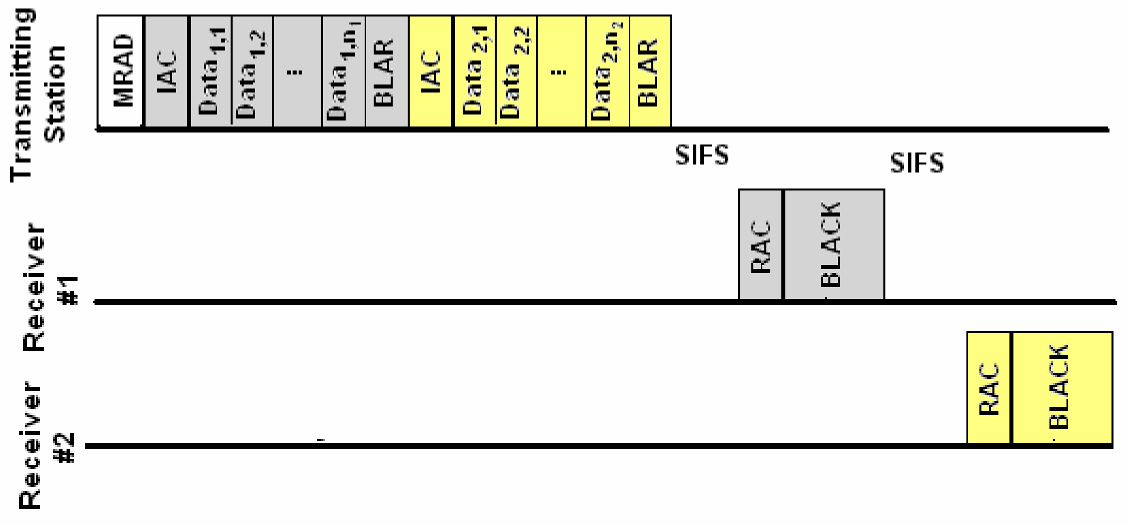


Figure 2.3 MRA Frame Exchange.

All data packets intended to different users in the MRA are sent at a common rate which is the lowest data rate of the data rate the receivers can support. Instead of rate adaptation via the IAC-RAC exchange, the common rate is determined by utilizing a database which indicates the average data rates used for transmitting to a user.

The extended version of MRA called Multiple-Receiver Multiple-Response Aggregation (MRMRA) enables the receiving stations to transmit any data packets which can be fit within a reverse duration granted by the initiator station. The amount of reverse time allowed for a receiver is indicated to the corresponding IAC packet.

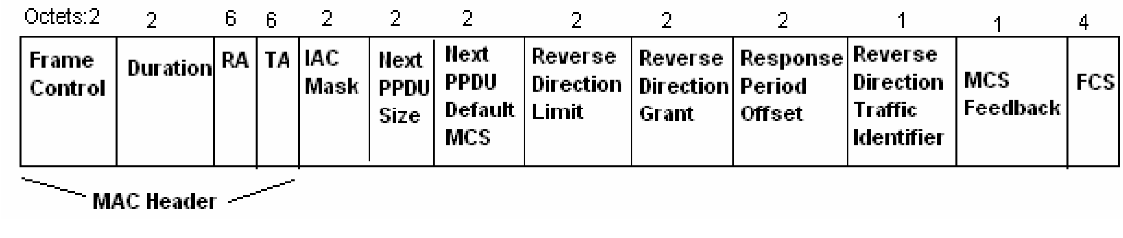
While forming an MRA, first the common rate is determined. Even though the determination of the common rate is not specified, main candidates are the highest rate available, the rate of the user with the longest queue or the rate of the user with the largest delay. When the MRA or MRMRA is constituted, there are parameters which must be satisfied. Typically the number of receivers must exceed a lower bound and the number of packets for any receiver has upper and lower limits. Note that the number of packets forming the MRMRA may also be limited by the TXOP limit.

2.2.1.2 Packet Formats

In 802.11n packet formats of previous 802.11 standards are extended to enable new features such as single- and multiple- destination aggregation. Control packets are transmitted at the basic rate.

Initiator Aggregation Control (IAC) Packet Format

The IAC packet is the extended version of RTS packets. The main difference between IAC and RTS is that IAC enables training and reverse direction traffic specifications. Transmission sequences are initiated by transmitting an IAC packet. The field stucture shown in Figure 2.4 is fixed, while some fields may not be used.



Octets:2	2	6	6	2	2	2	2	2	2	1	1	4
Frame Control	Duration	RA	TA	IAC Mask	Next PPDU Size	Next PPDU Default MCS	Reverse Direction Limit	Reverse Direction Grant	Response Period Offset	Reverse Direction Traffic Identifier	MCS Feedback	FCS

Figure 2.4 IAC Frame[1].

a. Duration: Time required for transmitting the aggregated packets, an RAC packet, a BLACK and BLAR plus three SIFS durations

b. IAC Mask: Bitmask indicating which IAC elements are present. Possible elements are RTS, TRQ(Training Request), MRQ(MCS Feedback Request),MFB(MCS

Feedback), FPD(Following Packet Desciptor), RDG(Reverse Direction Grant) and RDL(Reverse Direction Limit)

c. Next PPDU Size: Size in bytes of following PPDU that will be sent. Present if FPD is 1.

d. Next PPDU Default MCS: Default MCS without training information. Present if FPD is 1.

e. Reverse Direction Limit: Maximum time that can be granted for reverse transmission. Present if RDL is 1.

f.Reverse Direction Grant: Time available for reverse PPDU. Present if RDG is 1.

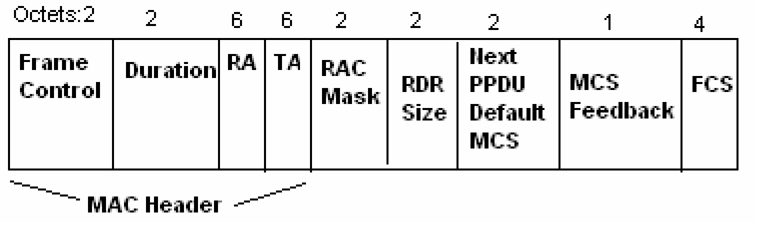
g.Response Period Offset(RPO): Separation between the IAC mpdu and the response PPDU. Should be at least SIFS. Present when RDG is 1.

h.Reverse Direction Traffic Identifier(RDTID): Indicates the AC or whether there is no constraint for reverse traffic. Present when RDG is indicated.

i.MCS Feedback: Recommended MCS value. Present when MFB is 1.

Responder Aggregation Control (RAC) Packet Format

The RAC packet is the extended version of CTS packets. RAC packets are transmitted as response to IAC packets. The field stucture shown in Figure 2.5 is fixed, while some fields may not be used.



Octets:2	2	6	6	2	2	2	1	4
Frame Control	Duration	RA	TA	RAC Mask	RDR Size	Next PPDU Default MCS	MCS Feedback	FCS

Figure 2.5 RAC Frame [1].

a. RAC Mask: Bitmask indicating which RAC elements are present. Possible elements are CTS, TRQ(Training Request), MRQ(MCS Feedback Request),MFB(MCS Feedback) and RDR(Reverse Direction Request)

b. RDR Size: Size in bytes of requested reverse direction flow. Present if RDR is 1.

c. Next PPDU Default MCS: Default MCS without training information. Present if RDR is 1.

d.MCS Feedback: Recommended MCS value. Present when MFB is 1.

Multiple Receiver Aggregate Desctiptor(MRAD) MPDU Packet Format

The MRAD packet is the initial packet sent at Multiple destination aggregation systems. The MRAD contains information about the receivers and it can be used for power savings.

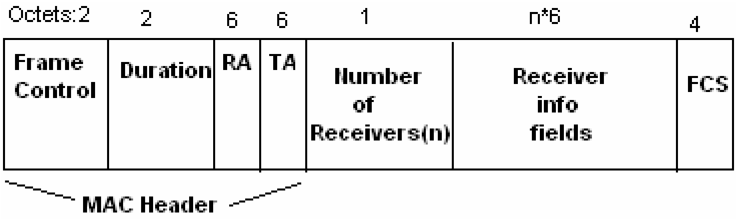


Figure 2.6 MRAD Frame [1].

The order of receivers in the receiver info fields are identical to the order of packets in the aggregate.

Aggregated PSDU Format

An aggregate Physical Layer service data unit (PSDU) is formed by a sequence of MPDUs. The MPDU consists of MPDU payload ,MPDU header and a FCS. Before each MPDU in the aggregation, there are MPDU delimiters.Padding octets are added in order to make it a multiple of 4 octets.

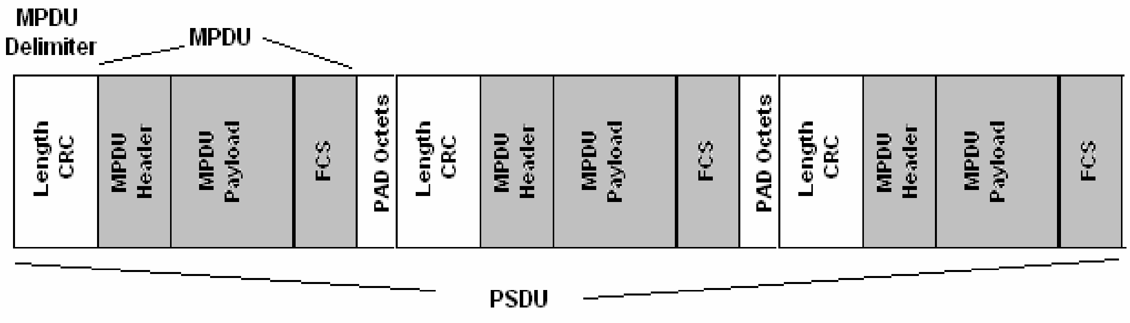


Figure 2.7 Aggregated PSDU Format [1].

The MPDU delimiter, formed by the fields ‘Reserved’, ‘MPDU length’, ‘CRC’, ‘Unique pattern’ is 32 bits long. MPDU delimiters are inserted to ease the recover of the aggregation structure when some of the MPDUs of the aggregate are errorenous.

Physical Layer Frame(PLCP) Format

The main aim of the PLCP header is to perform synchronization associated tasks. The seven tasks to be accomplished are Start of packet detection, Automatic Gain Control(AGC), Coarse and Fine Frequency Offset and Timing Offset estimations and channel estimation. The PLCP header proposed in 802.11n consists of two parts: legacy header and high-throughput(HT) header. The legacy header is designed for compatibility with 802.11a and 802.11g , while the HT header is specific to 802.11n. The structure can is presented in Fig. 2.8:

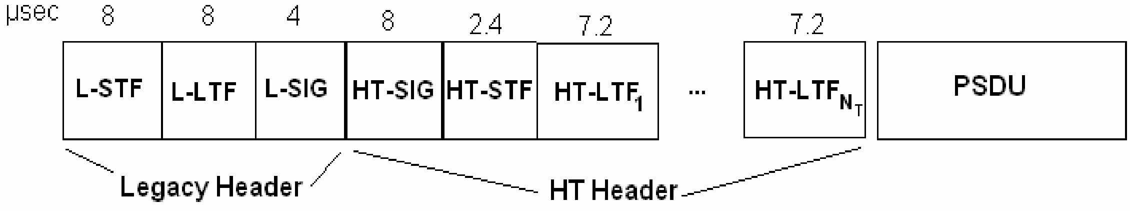


Figure 2.8 PPDU Format [1].

The legacy header is identical to the corresponding fields 802.11a/g. L-STF and L-LTF are legacy short- and long- training fields. The HT signal field carries information about the frame such as modulation and coding set, PPDU length and whether the packet is an aggregation. The HT-signal field is identified using a BPSK constellation with 90^0 phase offset compared with legacy signal field. In both legacy and HT training fields, the short training fields are responsible for AGC and coarse frequency and time offset estimation. On the other hand, long training fields are used for fine frequency offset estimation, fine time offset estimation and channel estimation. The number of HT-long training fields is the number of transmitter antennas.

2.2.2 Physical Layer Enhancements

2.2.2.1 Physical Medium Dependent(PMD) Structure

The physical layer of the 802.11n is MIMO extension over OFDM. Even though the initial proposal also defines 4x4 MIMO, the main contender for the standard is 2x2 MIMO. The general block diagram of the MIMO transmitter is shown is Figure 2.9

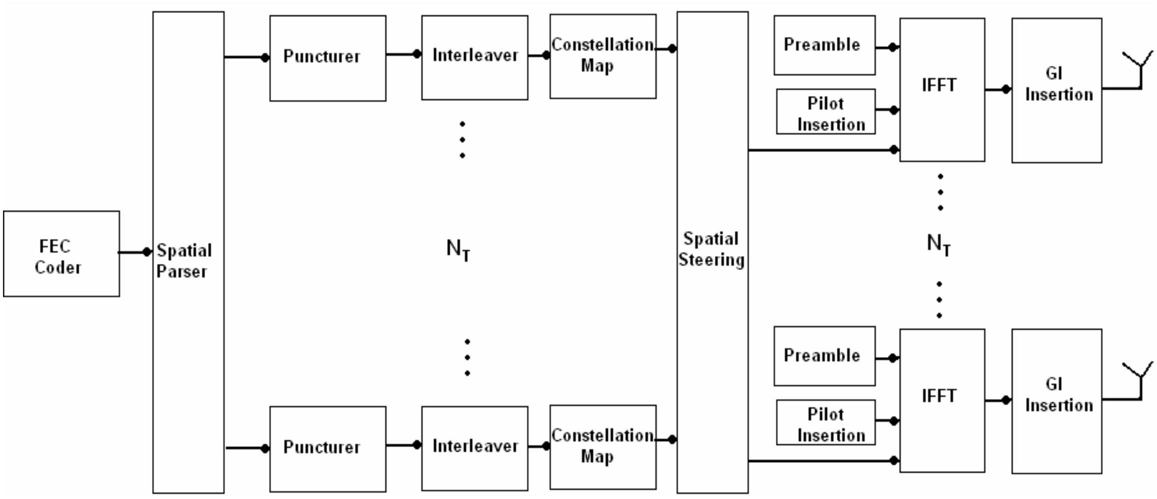


Figure 2.9 MIMO-OFDM Transmitter.

The data stream is first coded by a convolutional encoder. The coding rate can be varied adaptively according to the channel conditions if closed-loop MIMO is applied. The spatial parser demultiplexes the data into spatial streams for each antenna. Puncturer fixes the stream to the desired data rate. The coded streams are frequency interleaved and mapped to OFDM carriers. The spatial steering block is used in advanced beamforming systems where the each spatial stream can be assigned to any transmitter chain. Otherwise, each spatial stream is transmitted by a single transmitter chain and the spatial steering matrix can be replaced by an identity matrix. Finally, after IFFT, pilot tone and Guard Interval(GI) insertion, the signal is transmitted from the antenna.

2.2.2.2 MIMO Systems

In this Subsection the basic principles of Multiple-Input Multiple Output (MIMO) systems are presented. MIMO systems have emerged as the primary aid to increase data rates in wireless communication systems over the last decade[15]. As the name implies, in MIMO based systems both the transmitter and receiver ends of the communication system consists of multiple antennas. A view of a MIMO system with N_T transmit and N_R receive antennas can be seen in Figure 2.10:

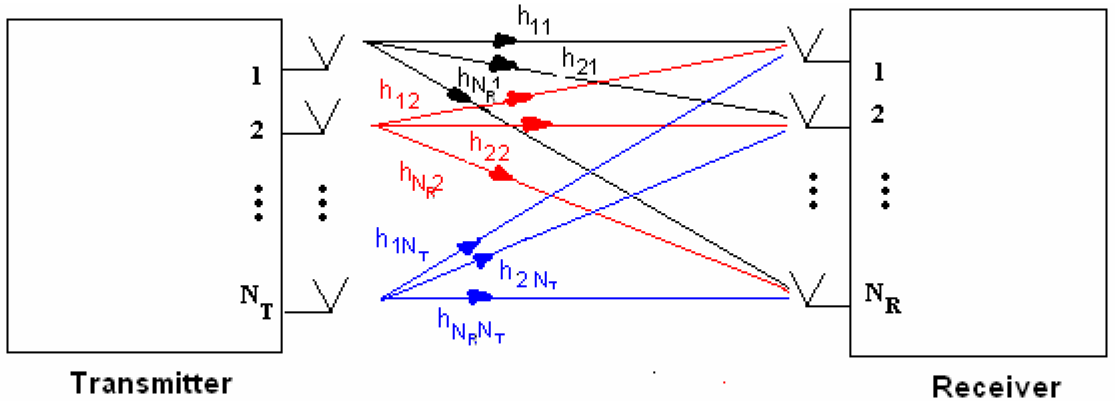


Figure 2.10 MIMO communication system with N_T transmit and N_R receive antennas.

All of the antennas transmit data at the same time and frequency band. In conventional communication systems without multiple antennas, the received signal can be represented in terms of the transmitted signal, a communication channel which is simply the point-to-point wireless link between the receiver and transmitter and additive noise. On the other hand, for MIMO systems, the signal received at a particular receiver antenna is a superposition of the signals transmitted from all transmit antennas [15]. The signals received at the receiver antennas are:

$$
\begin{aligned}
y_1 &= h_{11}s_1 + h_{12}s_2 + \cdots h_{1N_T}s_{N_T} + w \\
y_2 &= h_{21}s_1 + h_{22}s_2 + \cdots h_{2N_T}s_{N_T} + w \\
&\vdots \\
y_{N_R} &= h_{N_R1}s_1 + h_{N_R2}s_2 + \cdots h_{N_RN_T}s_{N_T} + w
\end{aligned}
\tag{2.1}
$$

where y_j is the signal received at the j'th receive antenna, s_j is the signal transmitted from the j'th transmit antenna, h_{ij} is the channel gain from transmit antenna j to receive antenna i and w is additive noise. MIMO systems are represented in a more compact form as follows:

$$\mathbf{y} = \mathbf{Hs} + \mathbf{w}, \tag{2.2}$$

where $\mathbf{y} = [y_1 y_2 \ldots y_{N_R}]$, $\mathbf{s} = [s_1 s_2 \cdots s_{N_T}]^T$ are received and transmitted signal vectors, $\mathbf{w} = [w_1 w_2 \cdots w_{N_T}]^T$ is the noise vector, and

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} \cdots & h_{1N_T} \\ h_{21} & h_{22} \cdots & h_{2N_T} \\ \vdots & \vdots & \vdots \\ h_{N_R1} & h_{N_R2} \cdots & h_{N_RN_T} \end{bmatrix} \tag{2.3}$$

is the N_T x N_R channel matrix. Channel capacity depends on the channel matrix as explained in the next subsection.

MIMO systems are mainly used in two modes: Spatial multiplexing and diversity. In spatial multiplexing, the data stream to be transmitted is demultiplexed and independent signals are transmitted over separate antennas. Spatial muliplexing leads to an increase in the total information transmitted. Provided that the individual channels are uncorrelated, the increase in data rate increases linearly with *min(N_T, N_R)*. On the other hand, in the spatial diversity mode of operation, the same signal is transmitted from all of the transmitter antennas. Spatial diveristy improves reliability of communication since the probability of suffering from deep fades from all subchannels is decreased. Discussions on the trade-off between spatial muliplexing and spatial diversity have been carried out in the literature [16].

Beamforming[15] can be applied to simplify receiver structure when the channel state information is availabe at the transmitter. From the principle of reciprocity, **H** is symmetric. If **H** is a square matrix and we take the Singular Value Decomposition(SVD) of **H** to have:

$$\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\mathbf{H}}, \tag{2.4}$$

where $\mathbf{\Lambda}$ is the diagonal matrix with the singular values of **H, U** is the right-singular vector of **H** and **V** is the left singular vector of **H.** **U** and **V** are unitary matrices. In beamforming **V** is available at the transmitter, $\mathbf{Vs}$ is transmitted instead of $\mathbf{s}$. At the receiver end, the received signal is multiplied by $\mathbf{U^H}$**.** After these operations, the received signal becomes

$$\mathbf{y} = \mathbf{U}^{\mathbf{H}}\mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\mathbf{H}}\mathbf{V}\mathbf{s} + \mathbf{U}^{\mathbf{H}}\mathbf{w}. \tag{2.5}$$

Since **U** and **V** are unitary matrices, i.e. $\mathbf{UU^H}$=$\mathbf{U^HU}$= $\mathbf{VV^H}$=$\mathbf{V^HV}$= **I**, we obtain

$$\mathbf{y} = \mathbf{\Lambda}\mathbf{s} + \mathbf{U}^{\mathbf{H}}\mathbf{w} \tag{2.6}$$

Eventually, the received signals at a specific antenna depends on the signal transmitted corresponding transmitter antenna. Thus, the subchannels are decoupled and the MIMO channel is decomposed into parallel subchannels with individual gains depending on the singular values of **H.**

2.2.2.3 MIMO and MIMO/OFDM Channel Capacity

The capacity for MIMO systems is calculated by evaluating the mutual information between the transmitted signal and both the received signal and channel matrix, and is given by the general well-known log-det capacity formula[1] [17]

$$C = \mathrm{B}\ \log_2(\det(\mathbf{I}_{\mathrm{M}} + \frac{\mathbf{H}\Sigma\mathbf{H}^H}{\sigma^2}))\ , \tag{2.7}$$

where M is the minimum number of receive (N_R) or transmit (N_T) antennas, and **H** is the N_R x N_T matrix of complex channel gains between the AP and the user, Σ is the transmit covariance matrix, σ^2 is the noise variance and $(.)^H$ denotes the Hermitian transpose operation.

In OFDM based systems, the channel capacity is calculated by partitioning the system into multiple subchannels that correspond to different subcarriers and accumulating subcarrier capacities. The channel matrix is an expanded version of the N_R x N_T by size N_C N_R x $N_C N_T$ with the original MIMO channel matrices over the diagonal axis and zero otherwise.

$$\mathbf{H} = diag\{\mathbf{H}(e^{j2\pi\frac{k}{N}})\}_{k=0}^{N-1} \tag{2.8}$$

Assuming equal number of transmit and receive antennas, M, the capacity formula for MIMO OFDM system is found as [19]

$$C = \frac{B}{N}\sum_{k=0}^{N-1} \log_2(\det(\mathbf{I}_{MN} + \rho\mathbf{H}(e^{j2\pi\frac{k}{N}})\mathbf{H}^H(e^{j2\pi\frac{k}{N}})))\ , \tag{2.9}$$

with

$$\mathbf{H}(e^{j2\pi\theta}) = \sum_{l=0}^{L-1}\mathbf{H}_l e^{-j2\pi l\theta}\ , \tag{2.10}$$

where N is the number of OFDM sub carriers, L is the number of resolvable paths and $\mathbf{H}_\mathbf{l}$ represents the l-the tap of the discrete-time MIMO fading channel impulse response [19].

[1] SISO Channel Capacity is given by $B\log_2(1+\rho|h|^2)$ [18]

2.3 Overview of Scheduling Algorithms for Wireless Networks

In a multiuser communication system with an infrastructure network topology, scheduling is the mechanism that determines which user should transmit/receive data in a given time interval. Scheduling is an essential element of the MAC layer due to its effect on the overall behavior of the system. Opportunistic scheduling algorithms maximize system throughput by making use of the channel variations and multiuser diversity. The main idea is to favor users that are experiencing the most desirable channel at each scheduling instant. In this subsection, we briefly present a survey of some of the existing scheduling disciplines for wireless networks.

Maximum Rate Scheduling (MRS)

In spatially greedy scheduling schemes, which are often denoted as Maximum Rate Scheduling (MRS), time varying channel variations are exploited. The selection metric is the channel capacity of the user. In other words, the scheduler allows the user with the best channel conditions to transmit at a given time instant [4]. Scheduling users according to the channel state can provide significant performance gain due to the independence of fading statistics across users. This phenomenon is called multiuser diversity. MRS method is shown to be optimal for capacity maximization. However, when WLAN throughput is considered, throughput maximization depends on frame sizes. In particular, with frame aggregation this algorithm may fail since specific stations are to be served more frequently. Their queues will not grow up, causing the aggregate sizes to be small, and small aggregate sizes results in low total throughput even though that user can operate at a high data rate. Another issue regarding the MRS algorithm is that the algorithm is prone to unfair throughput distribution between the users, since the users subject to poor channel conditions never get a chance to transmit. In MRS, the selected user $k*$ at the i^{th} transmission opportunity can be found as:

$$k_i^* = \arg\max_k C_i^k \quad , \tag{2.11}$$

where C_i^k denotes the channel capacity of the k^{th} user at the i^{th} transmission opportunity.

Proportional Fair Queuing (PFQ)

In Proportional Fair Queuing (PFQ) algorithm, the user with the best channel condition (capacity) relative to its own average capacity is selected [7]. The main aim of PFQ is to maximize throughput while satisfying fair resource allocation. If the users of all channels deviate from their mean capacities in similar ways, all users will gain access to the medium for similar proportions. Note that being selected similar proportions does not imply that the users have identical temporal share since transmission to users with low data rates take longer time durations for the same amount of data. In PFQ, the selected user k* can be found as:

$$k_i^* = \arg\max_k \frac{C_i^k}{\overline{C_i^k}}, \tag{2.12}$$

where $\overline{C_i^k}$ denotes the average channel capacity of the k^{th} user up to the i^{th} transmission opportunity.

Shortest Remaining Processing Time First (SRPT)

In [8], authors present Shortest Remaining Processing Time First (SRPT) method, where the metric is defined as the amount of time it takes to serve all the packets from a given queue. An opportunistic scheduler with this metric tries to choose the queue, which can be emptied in shortest amount of time. The selected user k* can be found as:

$$k_i^* = \arg\max_k \frac{Q_i^k}{C_i^k}, \tag{2.13}$$

where Q_i^k denotes the queue size of the k^{th} user at the i^{th} transmission opportunity.

Opportunistic Autorate (OAR)

In [9] [20], Opportunistic Autorate protocol(OAR), which is an opportunistic scheduler which takes into the effect of aggregation, users are served in a round-robin fashion. While serving each user, the number of packets transmitted for the user depends on the ratio of the user rate to basic rate, hence operating with larger aggregate sizes for users with better channel conditions. It is worthwhile to note that OAR provides temporal fairness since the packet transmission times for each user are equal.

Longest Queue (LQ)

Finally, we present a non-opportunistic scheduling scheme we denote as the Longest Queue (LQ) algorithm, which is also one of the considered schemes for 802.11n [1]. Using LQ, the scheduler simply selects the station with the largest number of packets in its queue. The channel states are not taken into account. In LQ, the selected user $k*$ is found as

$$k_i^* = \arg\max_k Q_i^k \tag{2.14}$$

The reasoning behind the LQ algorithm is to serve the user which is most likely to be served last and maximize the aggregate size at the same time, which leads to a higher throughput [1]. The queues of users which have not been served for a long time duration are likely to be long, increasing the scheduling metrics and eventually causing the assocaited user to be served.

CHAPTER 3

QUEUE AWARE OPPORTUNISTIC SCHEDULING

3.1 System Model

We consider the downlink of a MIMO wireless cellular system that consists of a single access point (AP) communicating with multiple mobile users. Fig. 3.1 depicts an example with two antennas. The system is a closed-loop MIMO OFDM system such that the mobile users measure their channel states and send them as feedback to the AP. As shown in Section 2.3, the knowledge of channel capacity is crucial in the design of schedulers.

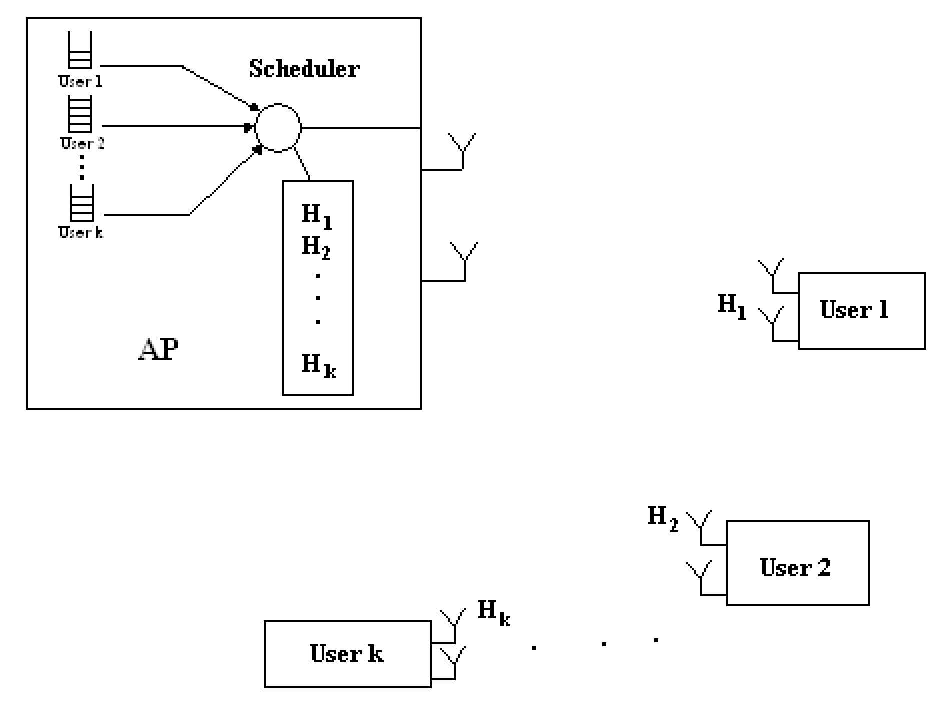


Figure 3.1 A 2x2 MIMO AP and mobile stations.

Considering the proposed 802.11n downlink transmission with contention-free aggregations described in Section 2.2.1.1, the maximum point-to-point throughput S_i for the i^{th} transmission opportunity is given by,

$$S_i = \frac{A_i.L_P}{\frac{L_{IAC}}{r_0} + \frac{L_{RAC}}{r_0} + 4.T_{PLCP} + DIFS + 4.\tau + 3.SIFS + \frac{L_{BLACK}}{r_0} + \frac{L_{BLAR}}{r_0} + \frac{A_i.(L_P + L_{MH})}{r_i}}. \quad (3.1)$$

A_i is the total aggregate size at transmission opportunity $i;$ $L_P,$ L_{IAC} , L_{RAC} ,$L_{BLACK},$ L_{BLAR} and L_{MH} are the length of the data, IAC, RAC, BLACK, BLAR packets and MAC header respectively in bits, T_{PLCP} is the duration of physical layer overhead, τ is the one way propagation delay, r_0 is the basic rate and r_i is the selected data rate during data transmission.

Observing (3.1), we see that there are mainly three adjustable parameters for throughput: the instantaneous data rates, the aggregate size and packet length, and all terms have a positive effect. Packet length is determined by the application, so we focus on the data rate and aggregate size terms. Without the overhead, aggregate size terms cancel out and throughput solely depends on the data rate used, justifying the usage of Maximum Rate Scheduling to maximize the throughput. However, the presence of overhead introduces the effect of aggregate size as well. Data rate is dependent on channel state and aggregate size is determined by queue size and the transmission opportunity limit.

Since throughput is significantly affected by channel states and queue states, the main aim of our work in this chapter is to enhance system throughput by discovering new scheduling metrics based on channel and queue states, while satisfying some level of fairness in throughput distribution. In sections 3.2, 3.3 and 3.4, we mainly propose four queue aware scheduling methods. The first three schedulers are based on the throughput expression and the fourth one is a heuristic based scheme. In Sections 3.5 we consider an extension of these schedulers to relaying.

3.2 Aggregate (Discrete)Opportunistic Scheduling(AOS/ADOS)

In Aggregate Opportunistic Scheduling (AOS) [21], we extend the Maximum Rate Scheduling (MRS) algorithm to situations with variable aggregate size. The aim of MRS is to select the user with maximum instantaneous channel capacity. Now, instead of instantaneous channel capacity, AOS selects the user that maximizes the system throughput, i.e., the selected user $k*$ at the i$^{\text{th}}$ transmission opportunity, is determined as

$$k_i^* = \arg\max_k S_i^k \,, \tag{3.2}$$

where S_i^k is the throughput of k^{th} user at i^{th} transmission opportunity, which is found as.

$$S_i^k = \frac{A_i^k \cdot L_P}{\frac{L_{IAC}}{r_0} + \frac{L_{RAC}}{r_0} + 4.T_{PLCP} + DIFS + 4.\tau + 3.SIFS + \frac{L_{BLACK}}{r_0} + \frac{L_{BLAR}}{r_0} + \frac{A_i^k.(L_P + L_{MH})}{C_i^k}} \cdot \tag{3.3}$$

Here C_i^k is the calculated capacity of user *k*, which depends on the channel state and A_i^k is the aggregate size for that user. Note that, $A_i^k = min(Q_i^k \,, A)$, where Q_i^k is the queue size of user *k* at *i*th opportunity and *A* is the maximum frame aggregate size, which is a system parameter.

Even though the metrics of each user are computed directly by inserting instantaneous capacity values, the actual data rates for transmission r_i are selected from a finite set of rates defined by 802.11n. In another scheduling algorithm, Aggregate Discrete Opportunistic Scheduling (ADOS), we propose that the AP first maps the instantaneous capacity values C_i^k to one of the 802.11n data rates from the set R_d={12,24,36,48,72,96,108,144,192,216} Mbps[1] to obtain r_i^k and then calculates the throughput values. Again, the selected user $k*$ at the i^{th} transmission opportunity, is found as

$$k_i^* = \arg\max_k S_i^k \qquad r_i^k = f(C_i^k), \quad r_i^k \in R_d \tag{3.4}$$

and

$$S_i^k = \frac{A_i^k \cdot L_P}{\frac{L_{IAC}}{r_0} + \frac{L_{RAC}}{r_0} + 4.T_{PLCP} + DIFS + 4.\tau + 3.SIFS + \frac{L_{BLACK}}{r_0} + \frac{L_{BLAR}}{r_0} + \frac{A_i^k.(L_P + L_{MH})}{r_i^k}}. \tag{3.5}$$

3.3 Proportional Aggregate Opportunistic Scheduling (P-AOS)

In addition to the AOS algorithm which maximizes (3.1), we also propose the extension of the Proportional Fair Queuing Algorithm (PFQ) applied to frame aggregation based transmission systems. Instead of favouring the user which maximizes the ratio of instantaneous-to-average channel capacity ratio as in (2.13), the user *k** which maximizes the ratio of instantaneous throughput to average throughput is selected at the i^{th} transmission opportunity, i.e.,

$$k_i^* = \arg\max_k \frac{S_i^k}{\overline{S_i^k}} \tag{3.6}$$

The main difference between the P-AOS algorithm and the PFQ algorithm from Section 2.3 is that the PFQ algorithm does not take into account aggregate size and performs scheduling decisions according to channel variations. Moreover, the throughput expression in (3.1) is more sensitive to the variations in aggregate size for users with high data rates compared with users operating under lower data rates since the transmission durations are shorter for the same aggregate size, causing the overhead terms to influence throughput more. This phenomenon can further cause P-AOS to prefer users with very good channel conditions.

3.4 Capacity Queue Scheduling (CQS)

Under typical opportunistic schedulers, users with high capacity links tend to have smaller queues as compared to users with low capacity links since they are selected more frequently. In the meantime, users subject to poor channel conditions are rarely served as they have low capacity links, which causes their queues to fill up. In order to design a scheduler that balances the channel and queue conditions, we represent the compromise between link capacity and queue size through a new metric, and call this method as Capacity Queue Scheduling (CQS) [21]. In this metric, both channel capacity and queue size have a positive effect on throughput. In CQS, the selected user k^* at the i^{th} transmission opportunity, can be found as:

$$k_i^* = \arg\max_k C_i^k Q_i^k \qquad (3.7)$$

The time varying nature of this metric leads to a fair allocation of resources. For instance, assuming initially that the queue sizes are similar in all users and one user's channel is superior with respect to others, that user will be selected and served, resulting in a smaller queue size in the next scheduling instant. Later, the advantage of this user with better channel characteristics will be reduced by its smaller queue size, and other users will have a chance to transmit. On the other hand, in Aggregate Opportunistic Scheduling the influence of a very large queue size on the selected user is limited since the scheduling metrics are upper bounded by the channel capacity. Nevertheless, load is a major factor which influences fairness performance since it directly affects the rate of increase of queue size.

3.5 Queue Aware Opportunistic Scheduling with Relaying

In this section, we try to take advantage of relaying in our schedulers through increased data rates due to reduced path loss. Relaying offers improvements in throughput, range extension in wireless networks, making use of multihop communication. With relaying, the main enhancement is that users do not directly transmit to the receiver and intermediate relaying stations are used to assist the transmission[22],[23]. Using intermediate relaying stations enables the communication to be carried out through shorter distances where the path loss much is lower as compared to direct transmission. The reduced path loss results in range extension, improved reliability (over the same range) which enables transmitters to use lower transmission powers or using higher data rates.

We consider only one relaying station. Our aim is to exploit relaying when it offers throughput enhancement with the information available at the AP. Figure 3.2 below shows the relaying scenario, with final station at R_f, relay station R_1 from the AP and the distance between relay and final station is R_2. and Figure 3.3 depicts the modified 802.11n transmission sequence with frame aggregation in relaying mode.

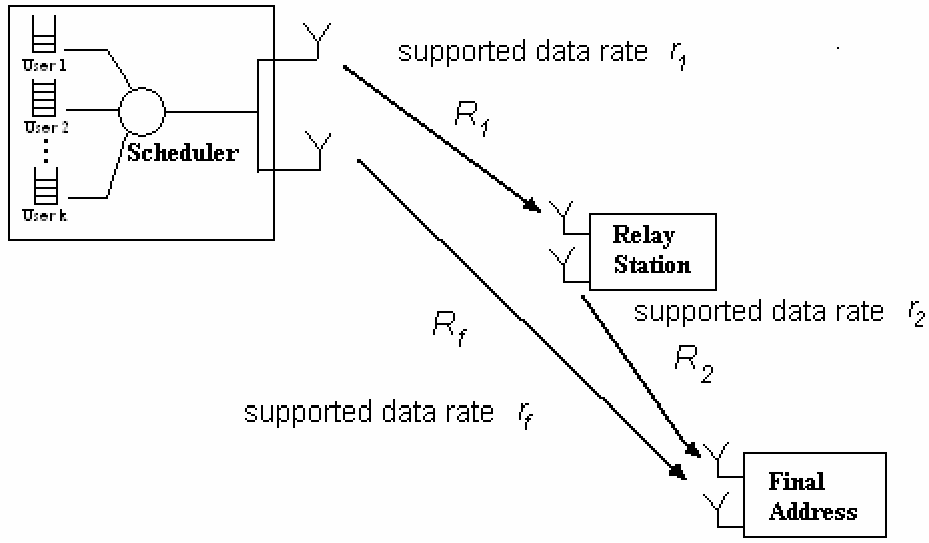

User 1
User 2
User k
Scheduler
supported data rate r_1
R_1
Relay Station
supported data rate r_2
R_2
R_f
supported data rate r_f
Final Address

Figure 3.2 Relaying with one relay station.

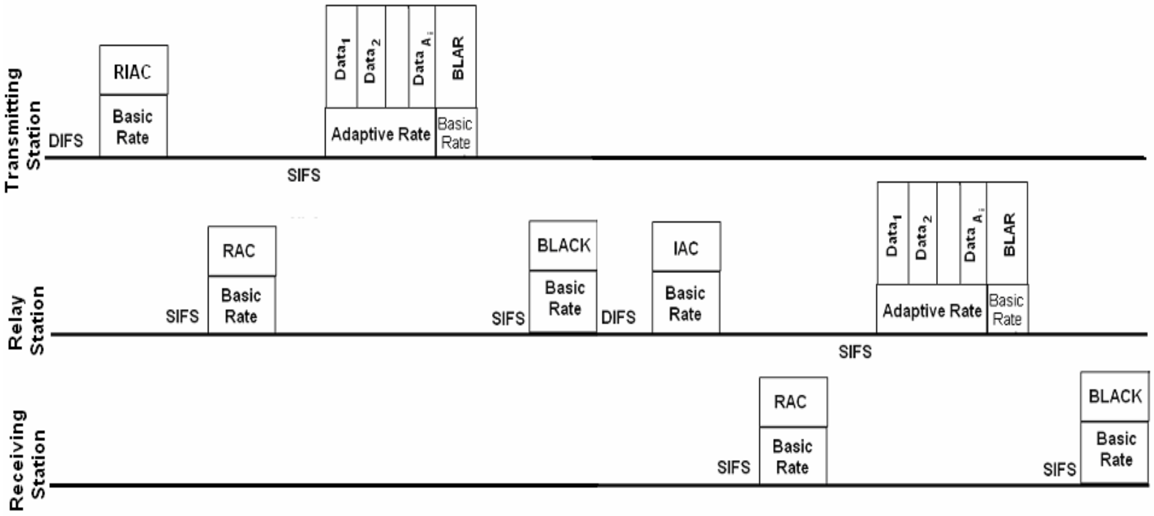


Figure 3.3 Frame aggregation with one relay station in IEEE 802.11n.

In order to implement relaying over the 802.11n protocol, the first IAC packet is modified as Relayed IAC(RIAC) adding fields to the packet which indicate whether relaying is required or not and the adddress of the relaying station. The relay station initiates another contention-free transmission sequence to the destination. A new transmission sequence will not be initiated at the AP unless the Block ACK is received from the final station. The principle of relaying structure can also be applied to mesh networks[24],[25] using the IEEE 802.11n.

In our problem we have an aggregated frame to be transmitted to a user and we aim to enhance the throughput by relaying. In order to determine whether relaying is beneficial for transmitting data to that user or not, we should compare the transmission duration with relaying to the situation without relaying. Throughout the analysis, we assume that the data rates supported from the AP to a user directly is r_f, from AP to the relaying station is r_1, and from the relaying station to the user is r_2 (See Figure 3.1). Without relaying, the total transmission duration is given by:

$$T_{direct} = \frac{L_{IAC}}{r_0} + \frac{L_{RAC}}{r_0} + 4.T_{PLCP} + DIFS + 4.\tau + 3.SIFS + \frac{L_{BLACK}}{r_0} + \frac{L_{BLAR}}{r_0} + \frac{A.L_P}{r_f} \quad (3.8)$$

$$\triangleq T_{overhead} + \frac{A.L_P}{r_f} \quad (3.9)$$

When relaying is employed, we transmit with data rates r_1 and r_2 , resulting in the transmission duration found as:

$$T_{relay} = 2.T_{overhead} + \frac{A.L_P}{r_1} + \frac{A.L_P}{r_2} \quad (3.10)$$

Relaying is beneficial if relaying offers a shorter transmission duration than direct transmission,i.e. when $T_{relay} < T_{direct}$.

An alternative approach to determine whether relaying is beneficial or not is to define an *equivalent relaying rate.* For this, we decompose (3.10) as follows:

$$T_{relay} = T_{overhead} + \frac{A.L_P}{r_1} + \frac{A.L_P}{r_2} + T_{overhead} \ , \tag{3.11}$$

We can re-write (3.11) as

$$T_{relay} = \left(\frac{1}{r_1} + \frac{1}{r_2} + \frac{T_{overhead}}{A.L_P} \right) A.L_P + T_{overhead} \ . \tag{3.12}$$

Note that the form of (3.11) is similar to (3.9) with the total duration as the sum of overhead delay and a rate-dependent term multiplied by the aggregated frame size, in bits ($A.L_P$). We define the equivalent relaying rate as the inverse of the rate-dependent term:

$$r_{equivalent} = \left(\frac{1}{r_1} + \frac{1}{r_2} + \frac{T_{overhead}}{A_i.L_P} \right)^{-1} \qquad \text{so that} \tag{3.13}$$

$$T_{relay} = T_{overhead} + \frac{A.L_P}{r_f} \ . \tag{3.14}$$

Note that $r_{equivalent}$ not only consists of rate-dependent terms, but it also depends on the aggregate size, A which in turn depends on the queue state for the final station. Increasing aggregate size increases the equivalent relaying rate.

Considering $r_{equivalent}$, our queue aware schedulers have been modified as follows:

i. For each destination station, calculate the effective relaying rate $r_{equivalent}$ using (3.13) for each possible intermediate station as a relay station.

ii. Select the best relaying station for that destination station, as the station which enables the maximum effective relaying rate.

iii. Compare the selected maximum effective relaying rate with the direct rate for the destination station.

iv. If the relaying rate is greater, prefer relaying while transmitting to that destination, with the relaying station specified in (ii).

v. Compute metric η_k, of the queue aware algorithm (AOS,…) using effective relaying rate if relaying is preferred for user *k*.

vi. Perform user scheduling by

$$k^* = \arg\max_k \eta_k \,. \tag{3.15}$$

If the relaying rate is smaller than direct rate, direct transmission is employed.

CHAPTER 4

SCHEDULING WITH CONTROLLED ACCESS

The algorithms presented in Chapter 3 perform user selection according to a scheduling metric, which considers instantaneous channel and queue states. For instance, in Aggregate Opportunistic Scheduling, the throughput is maximized for each transmission opportunity according to the instantaneous channel and queue states. However, when we consider overall throughput at a longer time scale, selecting that user that maximizes the throughput at the specific transmission opportunity may not be optimal since it may prevent transmitting with higher efficiencies in subsequent transmission opportunities. For each scheduling instant, even though the user which maximizes the scheduling metric is selected, by altering the scheduling mechanism by either introducing more transmissions or the order of transmissions to users with lower capacity users may enable the higher capacity users to transmit using higher aggregate sizes and better efficiency. However, whether the gain introduced by transmitting with higher efficiencies will prevail over increasing the ratio of poor channel users served should be analyzed.

In this chapter, our aim is to design scheduling algorithms which maximize the overall throughput over longer time intervals. In order to design such schedulers that make decisions over long time durations, estimating the evolution of the overall process by modelling the system using queueing theory is essential . In section 4.1, we first revise the bulk service model from queuing theory. Later on, we develop a queueing model of frame agregation in our systems, considering actual overhead and service rates. In Section 4.2 , we propose predictive schedulers that utilize the outcomes of this queuing model and control access proportions of users throughout the scheduling durations.

4.1 Analytical Modelling

4.1.1 Queuing Model of Bulk Transmission

We consider a the bulk service model [10] , where packets are served collectively in groups, i.e. bulks, and incoming packets are enqueued as shown in Figure 4.1 . When the service of a bulk is completed and the server becomes free , the next bulk is processed. Packets arrive in a Poisson fashion with average rate λ. All of the packets in the queue are served together if the number of packets is less than a limit bulk size , L. If the queue length exceeds L, only the first L packets are served. Service rate, μ is defined as the bulk service rate, which is assumed constant, regardless of the number of users in the bulk. The queueing system can be represented as shown in Figure 4.1 .

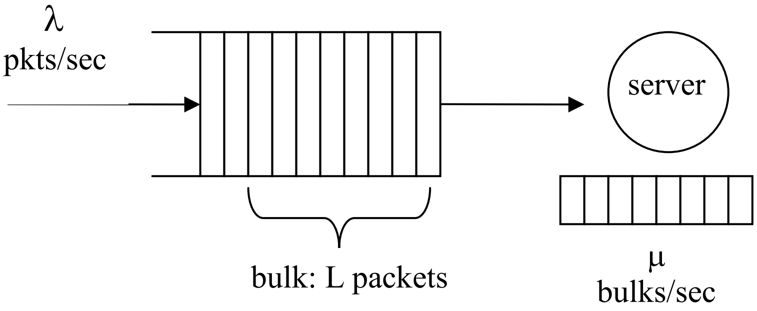


Figure 4.1 Bulk service system.

Defining the number of packets in the queue as the state of the queueing system, the Markov chain can be constructed as shown in Figure 4.2 . The arrival rate to a state is λ from one state to the incremented state and the bulk service rate is μ , where the state transition associated with bulk service corresponds to going back by L states:

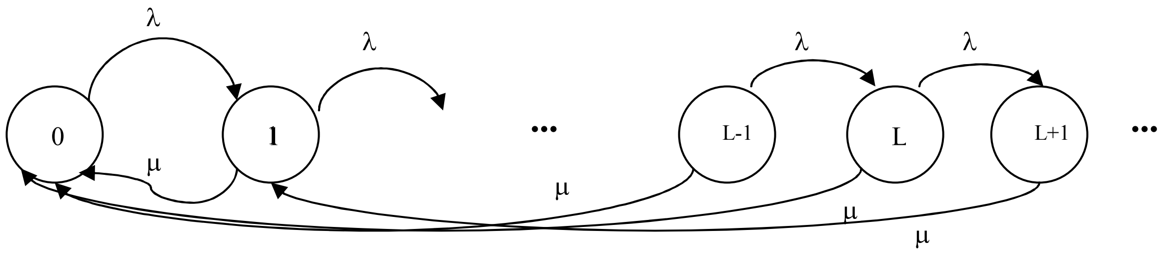


Figure 4.2 Markov chain representation of bulk service system.

Our aim is to find the state probabilites at steady state. The balance equations are easily seen to be:

$$\lambda p_0 = \mu p_1 + \mu p_2 + ... + \mu p_L \Rightarrow p_0 = (1/\lambda)\sum_{j=1}^{L} \mu p_j \tag{4.1a}$$

$$(\lambda + \mu) p_j = \mu p_{j+L} + \lambda p_{j-1} \qquad 1 \le j \tag{4.1b}$$

A widely used method to find the probability distributions of a queueing system is to use techniques to solve difference equations such as transforms[10]. The z-transform with positive exponent is defined as:

$$P(z) \triangleq \sum_{j=0}^{\infty} P_j z^j \tag{4.2}$$

Especially the following two identities for the alternative z-transform are used in the derivations throughout this chapter:

$$f_{n-k} \longleftrightarrow z^k F(z) \tag{4.3}$$

$$f_{n+k}, k > 0 \longleftrightarrow \frac{F(z)}{z^k} - \sum_{i=1}^{k} z^{i-k-1} f_{i-1} \tag{4.4}$$

Using identities (4.2) – (4.4), we can convert (4.1b) into the z-domain as:

$$(\lambda+\mu)[P(z)-p_0]=\frac{\mu}{z^L}[P(z)-\sum_{j=0}^{L}p_j z^j]+\lambda z P(z) \tag{4.5}$$

P(z) can be identified as

$$P(z)=\frac{\mu\sum_{k=0}^{L}p_j z^j-(\lambda+\mu)p_0 z^L}{\lambda z^{L+1}-(\lambda+\mu)z^L+\mu} \tag{4.6}$$

Although a term dependent on p_0 is required for complete representation of *P(z)*, we can simplify the related term using the first balance equation (4.1a) :

$$(\lambda+\mu)p_0 z^L=z^L(\mu\sum_{i=1}^{L}p_i+\mu p_0)=\mu\sum_{i=0}^{L}p_i z^L \tag{4.7}$$

P(z) can be expressed in a more compact form:

$$P(z)=\frac{\mu\sum_{k=0}^{L-1}p_j(z^j-z^L)}{\lambda z^{L+1}-(\lambda+\mu)z^L+\mu}=\frac{N(z)}{D(z)} \tag{4.8}$$

In order to take the inverse z-transform of *P(z)* to find an expression of p_j , the roots of the denominator *D(z)* are required . It can be shown that of the *L+1* roots of *D(z)* (and equivalently of the *L+1* zeros of *P(z)*) , one occurs at z=1, *L-1* are located inside the unit circle and one is located outside the unit circle[10]. Due to the fact that the z-transform of a probability distribution is analytical inside the unit circle, *P(z)* should be bounded, which implies that the *L-1* zeros of *P(z)* must also be the roots of the numerator *N(z)*. It is also obvious that one root of *N(z)* is at z=1. These facts enable us to simplify *P(z)* by cancelling out common factors of the form (z-z_k), where z_k are the common roots of *P(z)*. The remaining terms of *P(z)* are a constant divided by the term associated with the root out of the unit circle z_o:

$$P(z)=\frac{1}{K(1-z/z_0)} \tag{4.9}$$

Using the definition of *P(z)* and *P(1)*=1, *K* can be found as $\left(1-\frac{z}{z_0}\right)^{-1}$, leading to *P(z)* as:

$$P(z)=\frac{1-1/z_0}{1-z/z_0} \tag{4.10}$$

Finally, taking the inverse transform of *P(z)*, we achieve our goal of finding the steady state probabilities, p_j, which denote the distribution of packets in the bulk service system as[10]:

$$p_j=(1-\frac{1}{z_0})(\frac{1}{z_0})^j, \qquad \forall j \tag{4.11}$$

z_0 depends on *D(z)* and decreases with increasing λ/μ, leading to an increase in probability of being in higher states at high load.

4.1.2 Queueing Model of Proposed Aggregation Scheme

The bulk service model can be applied to systems employing frame aggregation where aggregate frame transmission is equivalent to bulk service and the maximum aggregate size is equivalent to L . However, in the bulk-service model introduced in Section 4.1, the bulk service rates are assumed to be constant independent of the queue state. This assumption implies that the service rate in terms of bits is directly proportional to the bulk size, since higher bulk sizes consist of longer frame sizes. Obviously, this assumption is not applicable in practical transmission scheduling including the schedulars we consider in 802.11n, since the channel data rate is constant, irrespective of the size of aggregation to be transmitted. In the presence of MAC and PHY overhead, the bulk service rate is further reduced since the actual data rates are not fully exploited.

Without MAC and PHY overhead, μ_j, the bulk service rate at the j^{th} state is inversely proportional to the number of packets in the aggregation, $A_j=\min(j,L)$, i.e.,

$$\mu_j = \begin{cases} \dfrac{\mu}{j}, & j < L \\ \dfrac{\mu}{L}, & j \geq L \end{cases} \tag{4.12}$$

When MAC and physical overhead is considered, the bulk service rate, μ_j ,at state j is computed as:

$$\mu_j = \begin{cases} \dfrac{\mu}{j} \cdot \left(\dfrac{j.L_P}{j.(L_P + L_{MH}) + L_{overhead} + r.T_{IFS}} \right) & j < L, \\ \dfrac{\mu}{L} \cdot \left(\dfrac{L.L_P}{L.(L_P + L_{MH}) + L_{overhead} + r.T_{IFS}} \right) & j \geq L, \end{cases} \tag{4.13}$$

where $L_{overhead}$ is the total length of control packets and the PLCP header ,T_{IFS} is the duration due to interframe spaces and r is the channel data rate. The data rate is adaptively selected according to the channel conditions. When there are more than L packets in the queue, the service rate remains constant due to the aggregation limit. The markov chain representation of aggregate frame transmission system is shown in Figure 4.3:

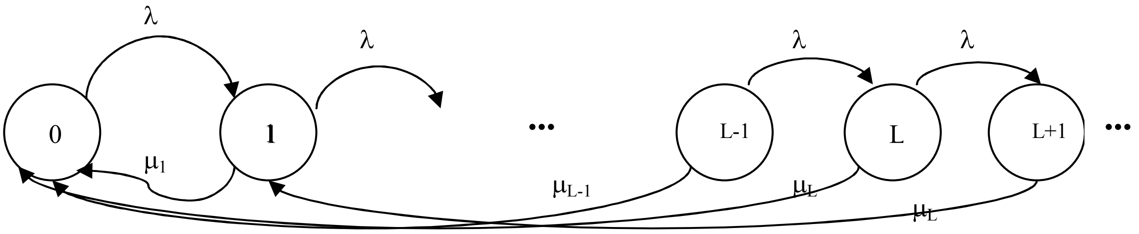


Figure 4.3 Markov-chain representation of frame aggregation system with overhead.

The balance equations of the above system can be obtained as follows:

$$\lambda p_0 = \mu_1 p_1 + \mu_2 p_2 + ... + \mu_L p_L \Rightarrow p_0 = (1/\lambda) \sum_{j=1}^{L} \mu_j p_j \tag{4.14a}$$

$$(\lambda + \mu_j)p_j = \mu_L p_{j+L} + \lambda p_{j-1} \qquad 1 \le j \le L \tag{4.14b}$$

$$(\lambda + \mu_L)p_j = \mu_L p_{j+L} + \lambda p_{j-1} \qquad j \ge L \tag{4.14c}$$

As in subsection 4.1.1, our aim is to find the stationary state probabilites p_j for a given load λ. Similar to the derivation for bulk service systems, we first evaluate the generator function P(z) related with the queueing system and take the inverse z-transform to achieve our goal. Converting the balance equations into the alternative form by taking the z-transform, we have

$$\lambda[P(z) - p_0] + \mu_L[P(z) - p_0] - \sum_{j=1}^{L-1}(\mu_L - \mu_j)p_j z^j = \frac{\mu_L}{z^L}[P(z) - \sum_{j=0}^{L} p_j z^j] + \lambda z P(z) \tag{4.15}$$

We can obtain P(z) as:

$$P(z)[\lambda + \mu_L - \frac{\mu_L}{z^L} - \lambda z] = p_0(\lambda + \mu_L) + \sum_{j=1}^{L-1}(\mu_L - \mu_j)p_j z^j - \frac{\mu_L}{z^L}\sum_{j=0}^{L} p_j z^j \tag{4.16}$$

$$P(z) = \frac{p_0(\lambda + \mu_L) + \sum_{j=1}^{L-1}(\mu_L - \mu_j)p_j z^j - \frac{\mu_L}{z^L}\sum_{j=0}^{L} p_j z^j}{\lambda + \mu_L - \frac{\mu_L}{z^L} - \lambda z} = \frac{N(Z)}{D(z)} \tag{4.17}$$

Even though the denominator of *P(z)* is identical to the bulk service system, the numerator *N(z)* is different so we focus on simplifying *N(z)*.

$$\frac{N(z)}{(-z^L)} = p_0(\lambda + \mu_L) + \sum_{j=1}^{L-1}(\mu_L - \mu_j)p_j z^j - \frac{\mu_L}{z^L}\sum_{j=0}^{L} p_j z^j \tag{4.18}$$

Combining the summation terms, we get:

$$\frac{N(z)}{(-z^L)} = p_0(\lambda + \mu_L) + \sum_{j=1}^{L}[(\mu_L - \mu_j)z^j - \frac{\mu_L}{z^L}z^j]p_j - \frac{\mu_L}{z^L}p_0 \tag{4.19}$$

Next, we insert p_o from (4.14a) and simplify:

$$\frac{N(z)}{(-z^L)} = \sum_{j=1}^{L}\mu_j p_j + \frac{\mu_L}{\lambda}\sum_{j=1}^{L}\mu_j p_j + \sum_{j=1}^{L}[(\mu_L-\mu_j)z^j - \frac{\mu_L}{z^L}z^j]p_j - \frac{\mu_L}{z^L}\frac{1}{\lambda}\sum_{j=1}^{L}\mu_j p_j \tag{4.20}$$

$$= \sum_{j=1}^{L} p_j[\mu_j + \frac{\mu_L\mu_j}{\lambda} + \mu_L z^j - \mu_j z^j - \frac{\mu_L}{z^L}z^j - \frac{\mu_L}{z^L}\frac{\mu_j}{\lambda}] \tag{4.21}$$

$$\Rightarrow \; N(z) = \sum_{j=1}^{L}[z^{L+j}(\mu_j-\mu_L) - z^L(\mu_j + \frac{\mu_L\mu_j}{\lambda}) + \mu_L z^j + \frac{\mu_L\mu_j}{\lambda}]p_j \tag{4.22}$$

Since z=1 is a root of the denominator of *P(z)*, *N(z)* should vanish at that root for a bounded *P(z)*. Therefore, *N(1)*=0, i.e.,

$$\sum_{j=1}^{L}[(\mu_j-\mu_L) - (\mu_j + \frac{\mu_L\mu_j}{\lambda}) + \mu_L + \frac{\mu_L\mu_j}{\lambda}]p_j = 0\,. \tag{4.23}$$

When $\mu_i = \mu_j = \mu_L = \mu$, which corresponds to the bulk service model, we have

$$N(z) = \sum_{j=1}^{L}[z^{L+j}(\mu-\mu) - z^L(\mu + \frac{\mu\mu}{\lambda}) + \mu z^j + \frac{\mu\mu}{\lambda}]p_j \sim \sum_{j=0}^{L-1} p_j\mu(z^j - z^L) \tag{4.24}$$

$$= \sum_{j=1}^{L}[\mu z^j - \mu z^L - \frac{\mu^2}{\lambda}z^L + \frac{\mu^2}{\lambda}]p_j \sim \sum_{j=0}^{L-1} p_j\mu(z^j - z^L) \tag{4.25}$$

Combining the summations and identifying the remaining terms, we verify that *N(z)* is consistent with the bulk service model as follows:

$$\sum_{j=1}^{L-1}[\frac{\mu^2}{\lambda} - \frac{\mu^2}{\lambda}z^L]p_j + \quad (\mu z^L - \mu z^L + \frac{\mu^2}{\lambda} - \frac{\mu^2}{\lambda}z^L)p_L \sim p_0\mu(1-z^L) \tag{4.26}$$

$$\sum_{j=1}^{L} p_j \frac{\mu^2}{\lambda}(1-z^L) \sim p_0\mu(1-z^L) = \frac{\mu}{\lambda}\sum_{j=1}^{L} p_j\mu(1-z^L) \tag{4.27}$$

Eventually the generator function *P(z)* of our frame aggregation model emerges as:

$$P(z) = \frac{\sum_{j=1}^{L}[z^{L+j}(\mu_j - \mu_L) - z^L(\mu_j + \frac{\mu_L\mu_j}{\lambda}) + \mu_L z^j + \frac{\mu_L\mu_j}{\lambda}]p_j}{\lambda z^{L+1} - (\lambda+\mu_L)z^L + \mu_L} \tag{4.28}$$

P(1)=1 should be satisfied as the global sum of probabilities is equal to 1, but since both *N(1)*=0 and *D(1)*=0, so we need to utilize the L'hospital rule. As such, we require that

$$\frac{N'(z)}{D'(z)} = 1 \tag{4.29}$$

Taking the derivatives of *N(z)* and *D(z)*:

$$N'(z) = \sum_{j=1}^{L}[(L+j)z^{L+j-1}(\mu_j - \mu_L) - Lz^{L-1}(\mu_j + \frac{\mu_L\mu_j}{\lambda}) + \mu_L j z^{j-1}]p_j \tag{4.30}$$

$$D'(z) = \lambda(L+1)z^L - (\lambda+\mu_L)Lz^{L-1} + 0 \tag{4.31}$$

Substituting z=1 we have

$$N'(1) = \sum_{j=1}^{L}[-L\mu_L + j\mu_j - L(\frac{\mu_L\mu_j}{\lambda})]p_j \tag{4.32}$$

$$D'(1) = \lambda(L+1) - (\lambda+\mu_L)L = \lambda - \mu_L L \tag{4.33}$$

and finally, *N'(1)=D'(1)*, i.e.,

$$\sum_{j=1}^{L}[-L\mu_L + j\mu_j - L(\frac{\mu_L \mu_j}{\lambda})]p_j = \lambda - \mu_L L\,. \tag{4.34}$$

The next step is to obtain state probabilities by taking the inverse transform of *P(z)*, which is not as simple as in the bulk service case. The fact that the bulk service rates are state-dependent has caused the order of *N(z)* to be greater than the order of *D(z)*, as a result we are not able to simplify the expression of *P(z)* as in (4.9), so we should take an alternative approach.

Similar to subsection 4.1.1, out of the L+1 roots of *D(z)*, L-1 of them are located within the unit circle, and from analyticity requirements of *P(z)*, *N(z)* must also vanish at each of these L-1 roots. This constraint results in a set of L-1 equations. Recalling that we also have the equation provided by (4.34), we obtain L equations that involve probabilities p_1 to p_L. We can also obtain p_0 from these probabilities via (4.14a). Hence the equation set is solved, resulting in the steady-state probabilities of the system up to state L. There is no closed form solution as in (4.11) for the bulk service model, so the equations are solved in MATLAB. We only have the steady state probabilities for states up to L, but this set will be sufficient for our scheduling purposes as discussed in Section 4.2.

An example of probability distribution of states up to L=63, and channel data rate r= 108 Mbps is shown in Figure 4.4. Figures 4.4.a) and b) depict state probabilities when incoming traffic load is 70 Mbps and 95 Mbps respectively. Load is expressed in terms of bits per second(bps), which is the product of packet load, λ (packets/second) and packet size(bits). As expected, the state probabilities for low states are much higher in the case of 70 Mbps load as compared to 95 Mbps.

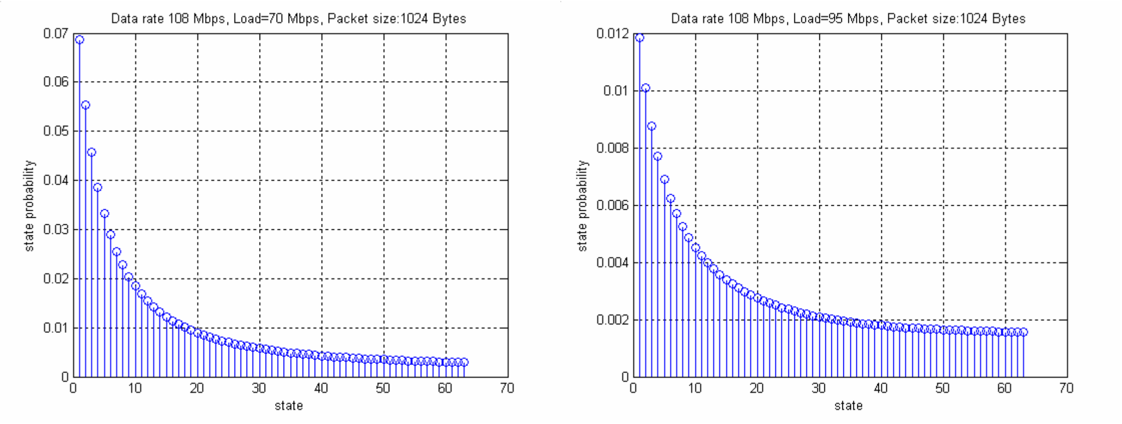


Figure 4.4 State probabilities up to L for load of a) 70 Mbps b)95 Mbps.

Next, in Figure 4.5, the effect of varying data rate on the state probabilities is shown. Incoming traffic load is 95 Mpbs and r=108 Mbps r= 216 Mbps in 4.5a and 4.5b respectively. Again, the state probabilities are concentrated towards the lower states for higher data rate case, since the queues are served faster.

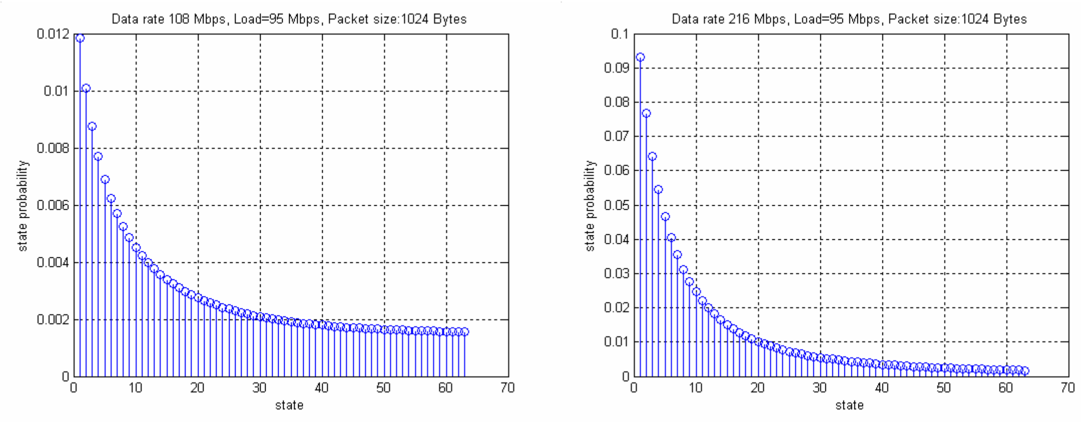


Figure 4.5 State probabilities up to L for data rate of a) 108 Mbps b)216 Mbps.

The analysis throughout this chapter up to now is valid when the maximum bulk service rate is larger than the load. If load exceeds the service rate, we assume that the state probabilities up to L are zero and the probability of exceeding L is 1. In Section 4.2, we make use of the state probabilities derived here in the design of predictive schedulers that define transmission sequences.

4.2 Predictive Scheduling with Controlled Access

In Section 4.1, we modeled the frame aggregation system with a data rate by taking overhead into consideration. We found the state probabilities up to the aggregation limit *L*. In this section, we propose to utilize these state probabilities in designing schedulers that maximize the total network throughput over a long time scale as opposed to the schedulers introduced in Chapter 3, which make decisions according to the instantaneous channel and queue states. The new schedulers control access proportions of individual users throughout the scheduling duration .Since throughput depends on the data rate and aggregate size, we find the expected aggregate size and expected throughput by averaging over calculated state probabilities.

Expected aggregate size can be directly obtained from the results of Section 4.1 . If the number of packets in the queue is less than *L*, then the aggregate size is the queue size. Otherwise, aggregate size is fixed by *L*. Hence, the expected aggregate size is given by

$$\overline{A} = \sum_{j=1}^{L} j.p_j + L.(1-\sum_{j=0}^{L} p_j) \qquad (4.35)$$

As an example, the variation of average aggregate size with load, served by a channel rate, *r*=108 Mbps and *L*= 63 is shown in Figure 4.6.

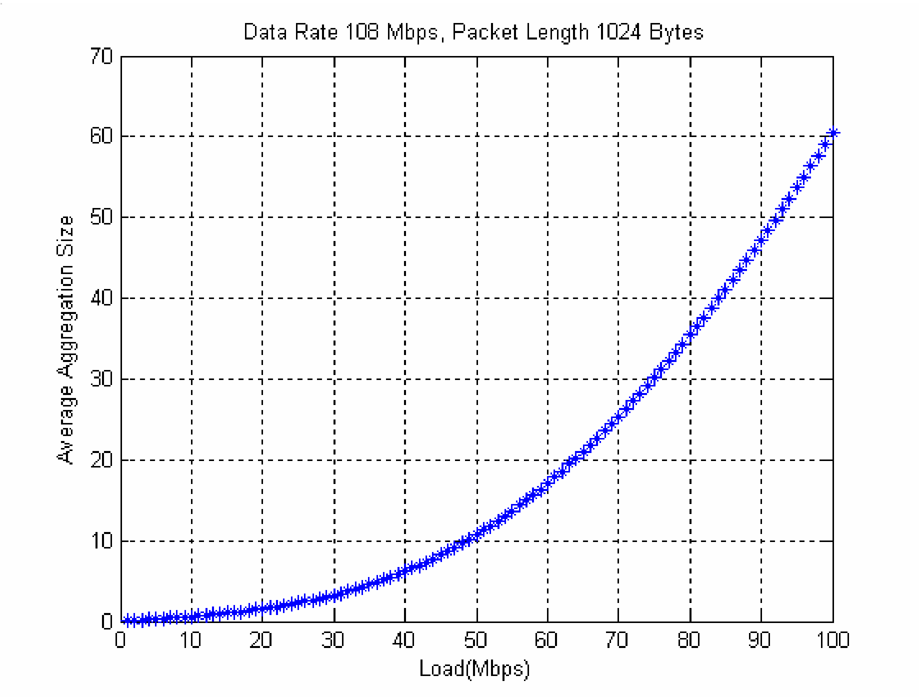


Figure 4.6 Average Aggregate size for varying Load at 108 Mbps Data Rate.

Throughput is a function of aggregation and the average throughput can be calculated by considering state probabilities of the aggregate size, as follows:

$$\overline{S}=\sum_{j=0}^{L} p_j S(A_j)+(1-\sum_{j=0}^{L} p_j)S(L) \tag{4.36}$$

where $S(A_j)$ is the throghput achieved with aggregate size A_j and is given by (3.1). It has been observed that throughput values have been equal to the load applied until the load exceeds the maximum service rate, which is the service rate for the maximum aggregate size. When the load exceeds the maximum service rate, throughput is saturated at that value.

$$\overline{S}=\begin{cases}\lambda_{bps} & ,\lambda_{bps}<S(L)\\ S(L), & \lambda_{bps}>S(L)\end{cases} \tag{4.37}$$

with $S(L)$ the throughput with aggregate size L and λ_{bps} is the load in terms of bps.

4.2.1 Predictive Scheduling with Access Guarantees(P-AG)

The aim of the scheduling algorithm derived in this section is to maximize total throughput in a multiuser network. However, as the name of the algorithm implies, the algorithm has a constraint that it should provide service guarantees for every user in the network. Scheduling is not applied over a single transmission sequence as in Chapter 3, but over multiple transmission sequences. The duration in which scheduling is applied is determined by the scheduling algoritm. Over this duration, the temporal access proportion of each user over all users is varied such that total system throughput is maximized. Our goal is to make an estimate for the aggregate size and throughput for each station by projecting the individual, per-user load level to the total load.

The previous analysis provides us with the average expected aggregate size and throughput for one user with a specific service capacity and load. Using this model, we can find the expected aggregate size and throughput by considering the related service and load parameters. The service rate depends on the instantaneous/average channel capacity and it is higher for users with high channel quality. In a multiuser scenario, considering downlink traffic, the AP can have packets destined to all users. For simplicity, we assume that load levels are identical for each destination station. In order to represent the multi-user system in our model, we modify the per-user load values such that the load values reflect the behaviour of the multi-user system. This modification is done by scaling the "*per-user load*" inversely by the temporal proportion π_n of user access to obtain the "*effective load*". If the temporal access proportion of a user is π_n, the actual service rate of the user is multiplied by π_n. From (4.17), it is seen that multiplying service rate by π_n with fixed load has the same effect as dividing load by π_n with fixed service rate. We prefer to divide load to obtain effective load since it is more feasible to calculate throughput and aggregate sizes for a finite set of data rates R_d with varying load than continuosly varying service rate for individual per-user load. For instance, the per-user load is 40 Mbps. If the access proportion of the user is 0.25, after scaling the effective load is found as 160 Mbps. The effect of access proportion is shown in Figures 4.7a and 4.7b.

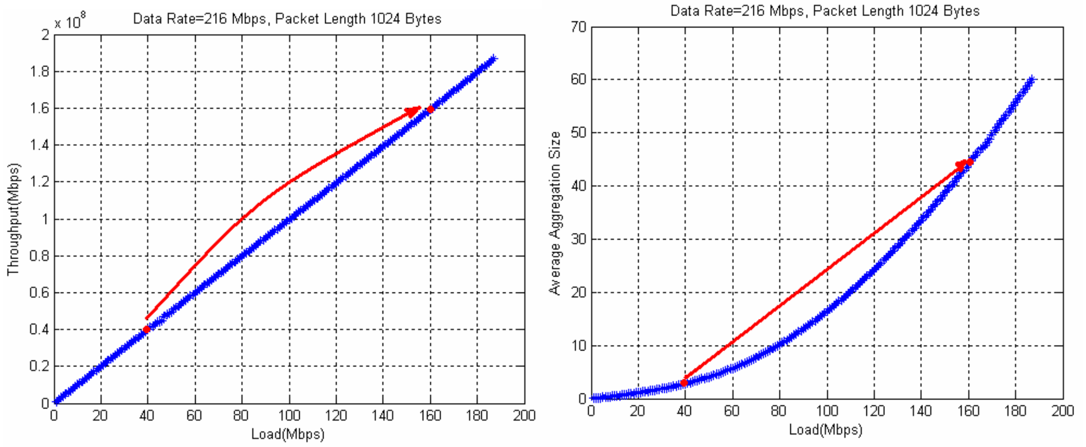


Figure 4.7 Effect of access proportion on a)Throughput b)Aggregate size.

The total network throughput depends on the weighted average of the individual throughput values of each user, as follows:

$$S_{total} \triangleq \sum_{n=1}^{N} \pi_n S_n \, , \tag{4.38}$$

where π_n is the temporal proportion of access for user n and S_n is the throughput while serving user n. A schematic is provided in Figure 4.8.

The proportion of access is a very critical issue. If the system is too “opportunistic”, good stations are served excessively, hence their access proportions will be high. This causes the effective load and the instantaneous throughputs for those stations to remain low. As the temporal proportion of good positioned stations is decreased, their effective load levels are increased, yielding an increase in their instantaneous throughputs. On the other hand, the proportion of low-quality stations is increased, leading to degradation in total throughput. A good balance must be maintained between these two factors. In fact, the problem can be redefined as maximizing (4.38) with the constraint.

$$\arg\max \sum_{n=1}^{N} \pi_n S_n \qquad \text{s.t.} \qquad \sum_{n=1}^{N} \pi_n = 1 \tag{4.39}$$

In order to maximize the total throughput, we perform an iterative search varying proportion values, and computing S_{total} , i.e., proportion vector $\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_N)$

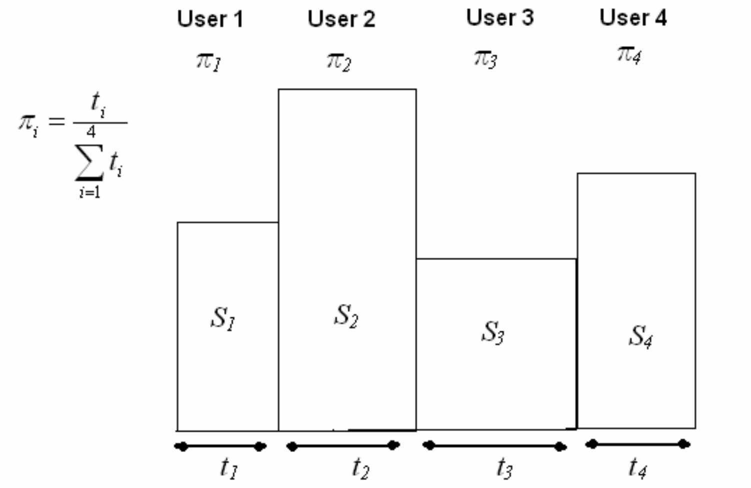


Figure 4.8 Transmitted sequence with asociated users.

In order to vary the proportions in general manner, we propose the following method :

i. First assign the initial proportions linearly, proportional with the average channel capacities, or the data rates r_n, i.e.,

$$\pi_n^0 = \frac{r_n}{\sum_{i=1}^{N} r_n} \qquad \forall n . \tag{4.40}$$

ii. For the i^{th} iteration of the search, take the exponents of the proportions (π_n^0) for all users

$$\pi_n^i = (\pi_n^0)^{\alpha_i}, \qquad \forall n . \tag{4.41}$$

where α_i is a parameter from a finite set associated with the i^{th} iteration. Since the summation of the temporal proportions of all users should be equal to unity, normalize proportions π_n^i .

$$\sum_{n=1}^{N} \pi_n^{i} = Q, \tag{4.42}$$

$$\pi_n^{i} = \frac{\pi_n^{i}}{Q}, \qquad \forall n. \tag{4.43}$$

After normalization, we have $\boldsymbol{\pi}^{i} = (\pi_1^{i}, \pi_2^{i}, ..., \pi_N^{i})$.

iii. Compute effective load values for each user,

$$\lambda_n^{i} = \frac{\lambda_n^{0}}{\pi_n^{i}}, \qquad \forall n \quad . \tag{4.44}$$

iv. Find $S^{i}(\lambda_n^{i})$ and $A^{i}(\lambda_n^{i})$ for $\forall n$ from the analytical model taking into account the data rates of each user.

v. Calculate $S_{total}^{i} = \sum_{n=1}^{N} \pi_n^{i} S_n^{i}$, record S_{total}^{i}.

vi. Change α_i, go to (ii)

Over all α_i values, the one maximizing the total throughput is selected. For that α_i, the temporal access proportions and aggregate sizes are determined of each user.

$$\max_{\alpha_i} S_{total}^{i} \Rightarrow \begin{cases} \boldsymbol{\pi}^{*} = [\pi_1^{*}, \pi_2^{*}, ..., \pi_N^{*}] \\ \mathbf{A}^{*} = [A_1^{*}, A_2^{*}, ..., A_N^{*}] \end{cases}. \tag{4.45}$$

Typically, for low α_i, the algorithm behaves similar to the LQ algorithm. As α_i is increased, throughput increases. However, the algorithm converges to the MRS algorithm when α_i is large. Users with high data rates are served with high access proportions leading to low effective loads and aggregation size, resulting in low throughput. Hence, the selected α_i result in behavior in between these two extreme cases. By determining the optimal access proportions of each user, we have obtained the transmission durations of each user relative to the total transmission sequnce in which scheduling is applied. The next step in scheduling is to realize a sequence of transmission over all users which satisfies the optimal proportions.

In order to define the transmission sequence, we assign each user a *turn number,* which indicates the number of times the user will be given access throughout the total scheduling duration. The turn number is determined in two steps; First the the ratio of the access proportion of each user to the transmission duration of serving that user once is found, i.e,

$$t_n^0 = \frac{\pi_n^*}{T_n} = \frac{\pi_n^*}{((A_n^*.L_P)/r_n + T_{overhead})} \quad , \tag{4.46}$$

where T_n is the transmission duration of serving user n once and $T_{overhead}$ refers to the overhead terms in (3.1). The transmission duration of serving a user once is longer when the data rate supported for that user is low or aggregate size is large. Hence, (4.46) will cause the turn number to decrease for such users. Accordingly, the turn number of users with poor channel conditions is expected to be lower as compared to users with better channel conditions.

Next, the lowest round number is determined, and the turn number associated with that station is set to 1. Thus, the station with poorest channel conditions is guaranteed service throughout the the total transmission sequence defined by the algorithm. The round numbers of other stations are scaled with respect to that round number. Rounding is performed when necessary since the rations may not result in integers.

$$t' = \min\left(\frac{\pi_1^*}{T_1}, \frac{\pi_2^*}{T_2}, \ldots, \frac{\pi_N^*}{T_N}\right), \tag{4.47}$$

$$t_1 = \frac{t_1^0}{t'}, t_2 = \frac{t_2^0}{t'}, \ldots, t_N = \frac{t_N^0}{t'} . \tag{4.48}$$

After determining the turn numbers of each user, the transmission sequence is determined. Users are sorted with ascending order of turn numbers and scheduling is performed by assigning transmission starting with smallest round numbers. Figure 4.9 illustrates scheduling for a network of five stations with associated turn numbers in Table 4.1 :

Table 4.1 Example turn numbers

User	a	b	c	d	E
Turn Number	1	1	2	2	3

STA_a^1	STA_b^1	STA_c^1	STA_c^2	STA_d^1	STA_d^2	STA_e^1	STA_e^2	STA_e^3

Figure 4.9 An example transmission sequence.

The degree of opportunism increses with increasing α_i. Yet, this algorithm offers bandwith(QoS) guarantees since every station is eventually served, which may not be valid for methods like AOS and MRS. The exponent α_i determines the degree of fairness, which varies with load. When load is low, the exponent is relatively low since the aggregate sizes for high-quality stations should be kept high to provide satisfactory throughput values, which can be done by reducing the access proportion. On the other hand, as the load is increased, the stations can transmit with higher aggregate sizes without needing to decrease their access proportion, so α_i is increased.

4.2.2 Predictive Scheduling with Time-domain Waterfilling(P-WF)

In section 4.2.1, throughput is maximized over a scheduling duration with each user guaranteed access throughout the scheduled period. Optimal temporal proportions of access for each user are determined after carrying search. In this section, the aim is to maximize throughput by applying the principle of waterfilling. The principle of waterfilling is commonly used in the field of information theory, especially in problems concerning power control [18]. In a multiple-channel scenario, the main idea behind the waterfilling principle is to allocate more power to transmissions over higher capacity channels and lower power to transmissions over lower capacity channels. In waterfilling problems, the aim is to maximize a weighted average with a constraint. An example formulation is to find the optimal (*x_1 ,x_2, ... x_N*) in order to

$$\max \sum_{i=1}^{N} (\beta + \gamma_i x_i) \text{ with the constraint } \sum_{i=1}^{N} x_i = 1 \qquad (4.49)$$

Utilizing Lagrangian methods, the waterfilling solution to the problem is given as

$$x_i^{opt} = (\mu - \frac{\beta}{\gamma_i})_+ \, , \; i = 1,...,N \, , \qquad (4.50)$$

where *$(\theta)_+$* denotes *$max(\theta,0)$*. The allocation with some modes unused are depicted in Figure 4.10. In waterfilling applications, usually the maximized quantity is capacity and the allocated x_i is power. Hence, throughout the text we assume that the (4.49) and (4.50) correspond to such a case.

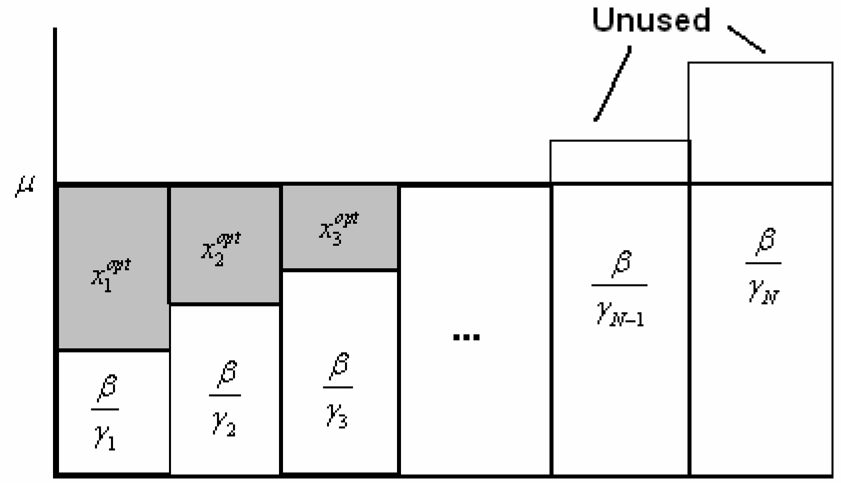


Figure 4.10 Schematic of Waterfilling Algorithm

In this work, our aim is to maximize the overall throughput over a sequence of transmissions via scheduling. The throughput over the transmission sequence is given by (4.51).

$$S_{total} \triangleq \sum_{n=1}^{N} \pi_n S_n \qquad \text{s.t.} \qquad \sum_{n=1}^{N} \pi_n = 1 \tag{4.51}$$

We apply the concept of waterfilling to the time proportions π_n to find the proportions that maximizes (4.51) . We call this method as "*time-domain waterfilling*". In contrast to the predictive scheduler introduced in Section 4.2.1 , here we do not force each station to have access opportunity within each transmission sequence; hence users with poor channel conditions may not be served (This fact is consistent with waterfilling schemes where poor channel states are not allocated power if their Signal to Noise Ratio (SNR) fall below the cutoff value μ). By comparing (4.51) with (4.49), we can exploit the mathermatical analogy between these equations. Here, total throughput is analogous to total capacity and power constraint is analogous to access proportion constraint. On the other hand, even though both the power proportions and time proportions are weighting factors, (4.49) also includes an additive term which is crucial for the remaining of the waterfilling analysis. In order to achieve a full analogy between the equation pairs, we add a constant to each term in the summation of (4.51) :

$$S' \triangleq \sum_{n=1}^{N} (\alpha + \pi_n S_n^i), \tag{4.52}$$

Maximizing *S'* is equivalent to maximizing S_{total} , so our problem becomes

$$\arg\max \sum_{n=1}^{N} (\alpha + \pi_n S_n^i) \qquad \text{s.t.} \qquad \sum_{n=1}^{N} \pi_n = 1. \tag{4.53}$$

The waterfilling solution is given as:

$$\pi_n = \left(\varsigma - \frac{\alpha}{S_n} \right)_+ \tag{4.54}$$

We can not compute π_n values directly, since S_n depends on π_n. In order to overcome this coupling in the waterfilling terms, we apply an iterative procedure to π_n values.

According to our queueing model, we can express S_n in terms of the per-user load λ_n, time access proportions π_n, supported data rate r_n and MAC overhead as follows:

$$S_n = f(\pi_n) = \begin{cases} \dfrac{\lambda_n}{\pi_n} \ , \ \dfrac{\lambda_n}{\pi_n} < S(L) \\ \\ S(L), \dfrac{\lambda_n}{\pi_n} > S(L) \end{cases} , \tag{4.55}$$

where *S(L)* is the maximum throughput offered by using data rate r_n , which is equal to the throughput that can be achived with the highest aggregate size, *L* allowed.

Our iterative time-domain waterfilling algorithm can be described as follows:

i. π_n^0 is initialized as 1/*N* for n=1...*N*.

ii. For iteration i=1,2,..*I*, access proportions are calculated using the formula:

$$\pi_n^{i+1} = \left(\varsigma - \frac{\alpha}{f(\pi_n^i)} \right)_+ \tag{4.56}$$

The threshold is also evaluated for each iteration, using the constraint

$$\sum_{i=0}^{N} \left(\varsigma - \frac{\alpha}{f(\pi_n^i)} \right)_+ = 1 \tag{4.57}$$

The equation is solved assuming all of the access proportions are greater than zero:

$$\varsigma = \frac{1}{N} + \frac{1}{N} \sum_{i=0}^{N} \left(\frac{\alpha}{f(\pi_n^i)} \right) \tag{4.58}$$

After evaluating ζ, it is checked whether $\dfrac{\alpha}{f(\pi_n^i)} > \varsigma$ is satisfied for all users. If that condition is satisfied, the iteration is completed. Otherwise, the cutoff is calculated by eliminating users with low S_n^i until the number of users surpassing the cutoff is consistent with the number of terms in (4.58).

After a finite number of iterations, the access proportions π_n converge and are determined. According to π_n, the transmission sequence is constructed after calculating round numbers and selecting aggregate sizes as given by our queueing model as explained in Section 4.2.1.

CHAPTER 5

SIMULATION MODEL

We have evaluated the performance of the aforementioned scheduling disciplines via simulations. The simulations are carried out in the OPNET (Optimized Network Engineering Tool) [26] simulation environment, which models the wireless channel and physical layer as well as 802.11 MAC layer with 802.11n enhancements and scheduling algorithms.

OPNET is an event driven simulation program. Networks are modeled in three hierarchies: network, node and process level. The network level defines the types of objects, and the locations of the object, which affects the topology. The medium level is the node level which consists of modules which usually correspond to the layers of the OSI architecture and traffic sources and sinks where packets are generated according to the specified load pattern and received. The lowest level hierarchy is the process level where the MAC specific protocols, physical layer functions and are modeled. Functions defined in the process models are supplemented by external codes. Physical medium such as channel model are modeled by the aid of pipeline stages which define attributes such as path loss, fading, background noise.

Figure 5.1 shows an example network model and the node level. The node model consists of a source, sink transcievers. Packets from the source pass through the MAC and PHY modules before transmission. A received packet is processed in thge PHY and MAC modules before being destroyed in the sink. The MAC and PHY modules, together with external models and pipeline stages model the related algorithms.

For the fading model, the Channel B model [27] developed by TGnSync group is implemented. In this model for small office environments and non line-of-sight (NLOS) conditions are modeled. In the physical layer, a 2x2 MIMO antenna configuration is assumed. OFDM parameters such as guard interval, number of subcarriers etc. are chosen according to the 802.11n specifications in [27]. IEEE 802.11n data rates are adaptively selected from the set {12,24,36,48,72,96,108,144,192,216}Mbps according to the instantaneous channel conditions.

The basic rate, i.e. the common rate for control packet transmission has been selected as 24 Mbps, since that was the lowest rate used by any of the stations for our topologies through sections 6.1 to 6.4. Some of the MAC related parameters of the simulation model are given in Table I. The maximum number of packets allowed in frame aggregation is assumed as 63.

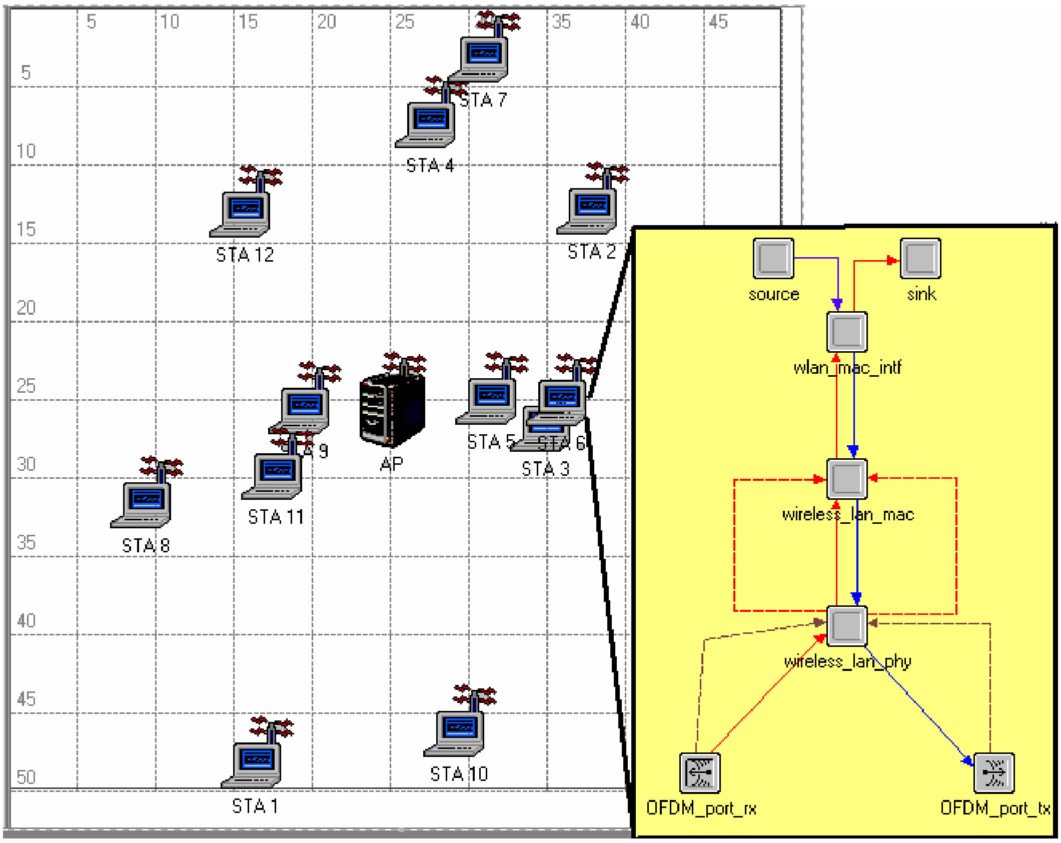


Figure 5.1. Network and Node Level in OPNET.

5.1 MAC Layer

The process model of the MAC module is shown in Figure 5.2. The process model consists of many states, which resembles a finite state machine. The simulations are event-driven by interrupts and according to the interrupts in hand transitions between states occur. In OPNET, there are two types of interrups: stream interrupts or self interrupts. Stream interrupts are due to changes in the status of higher layers, e.g. a packet arrival from upper layers to be transmitted, or lower layers, e.g. a packet that has been received. Self interrupts are due to the internal operation of the MAC protocol and are mainly related with the end of interframe spaces or backoff slots and frame timeouts.

INIT and BSS_INIT states are responsible for parameter initialization. The IDLE state is the default state without any tasks to execute. When a packet should be transmitted, depending on the medium status from carrier sense, either the TRANSMIT or DEFER state is entered. BACKOFF state is entered to perform the exponential backoff algorithm. All frames, whether the frames are control or data, are transmitted in the TRANSMIT state. After transmission, the process waits for responses if an aggregate was sent or response if a single packet was sent. The FRM_END state sets timers which defines time bounds for the acknowledgement reception. If reverse transmission is allowed the state WAIT FOR REVERSE is entered. The responding station which is granted reverse causes the flow to enter the TRANSMIT state back from REVERSE_TURN_POINT. If the station has data to transmit after the current transmission sequence is finished, execution enters to the DEFER state. Otherwise, the IDLE state is entered.

Table 5.1 Some MAC related parameters

Parameter	Value
τ	1 μ sec= 10^{-6} sec.
SIFS	16 μ sec= 16 X 10^{-6} sec.
DIFS	34 μ sec= 34 X 10^{-6} sec.
Basic bit rate	24 Mbps.
PLCP overhead	44.8 μ sec= 448 X 10^{-7} sec.
T_{IAC}	11.2 μ sec = 112 X 10^{-7} sec.
T_{RAC}	8.7 μ sec = 87 X 10^{-7} sec.
T_{BLACK}	48.7 μ sec = 487 X 10^{-7} sec.
T_{BLAR}	9 μ sec = 9 X 10^{-7} sec
L_{MH}	272 Bits

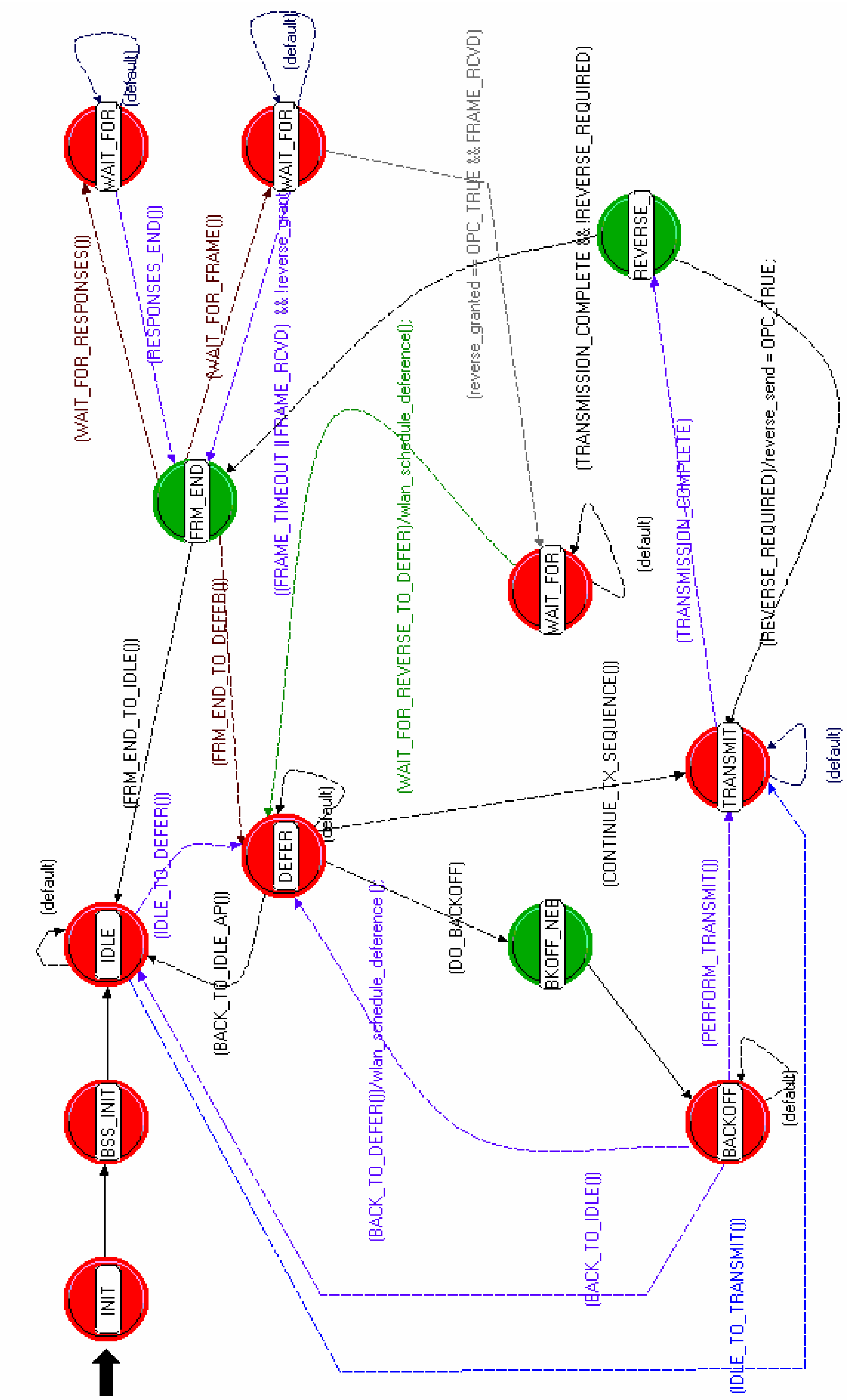


Figure 5.2 Process model of 802.11n MAC in OPNET.

5.1.1 Frame Aggregation

In the 802.11n MAC process model, the type of transmission sequence and the adress(es) of the receiving stations are determined in the **wlan_define_tx_type()** function. This fuction also makes use of the scheduling function explained in 5.1.2. According to the transmisison mode selected, there are three main funtions related with frame aggregation . The **wlan_burst()** function performs single destination aggregation. If the mode is closed loop, a TRAINED BURST is formed where available packets to the receiving station are formed together with IAC and BLAR packets. The number of packets that are included in the burst may also be limited due to the TXOP limit thus the size of the burst is determined by calculating the time left from the allowed TXOP after overhead and control packet durations are subtracted. **wlan_reverse_burst()** is responsible for the transmission of reverse aggregation if reverse transmission is allowed and the duration is enough for aggregation. If the transmission sequence decided is a MRA or MRMRA , the **wlan_multi_dst_burst()** function is called. The **wlan_multi_dst_burst()** function forms the aggregation according to the multi_dst_list created by the **wlan_can_form_mrmra()** function.

5.1.2 Scheduling

The scheduling funcition is implemented by the function **wlan_select_transmission_type()**. This function originally selects the user to be scheduled according to the Longest Queue(LQ) algorithm and determines if aggregation will be applied according to the queue state, but the function has been modified such that it determines the selected station by also considering parameters such channel capacity associated with each user. The load-dependent controlled access schedulers in Section 4.2 have also been implemented by modifying this function. First, the proportions of each user are determined, followed by round numbers accordingly. Finally, the users to be served and the corresponding aggregate sizes are determined according to the transmission sequence order.

5.2 MAC- PHY Interface

The MAC-PHY interface consists of two functions from each module. The **wlan_physical_layer_data_arrival()** function is in the MAC module which processes packets arrived from the physical layer. If the address of the packet is the receiving station, the packets are accepted and the behaviour is adjusted according protocol requirements. The **wlan_mac_data_arrival()** function in the physical layer sets the required fields and defines the required coding and modulation.

5.3 Physical Layer

The physical layer module is responsible for the transmission of packets to the wireless medium. The main functions performed are modifying packets coming from the MAC module, processing received packets from the receiver and extracting feedback infromation from the incoming packet.

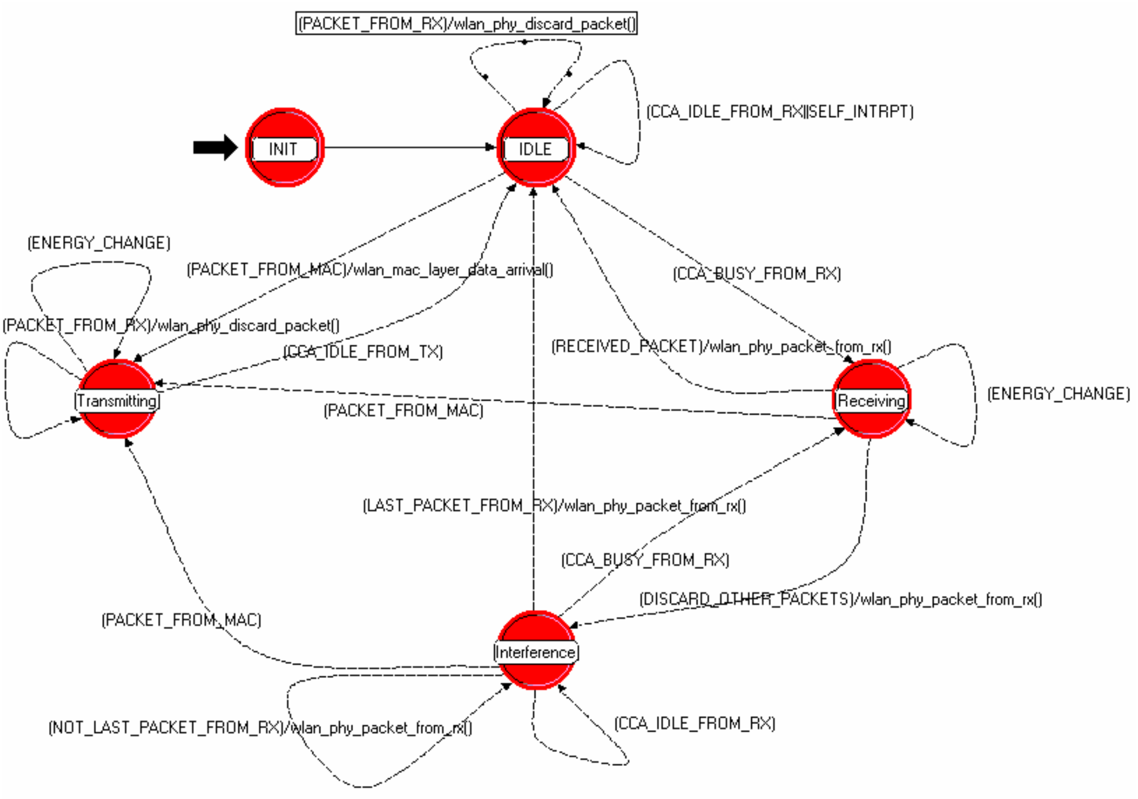

(PACKET_FROM_RX)/wlan_phy_discard_packet()
(CCA_IDLE_FROM_RX||SELF_INTRPT)
INIT
IDLE
(ENERGY_CHANGE)
(PACKET_FROM_MAC)/wlan_mac_layer_data_arrival()
(CCA_BUSY_FROM_RX)
(PACKET_FROM_RX)/wlan_phy_discard_packet()
(CCA_IDLE_FROM_TX)
(RECEIVED_PACKET)/wlan_phy_packet_from_rx()
Transmitting
(PACKET_FROM_MAC)
Receiving
(ENERGY_CHANGE)
(LAST_PACKET_FROM_RX)/wlan_phy_packet_from_rx()
(CCA_BUSY_FROM_RX)
(DISCARD_OTHER_PACKETS)/wlan_phy_packet_from_rx()
(PACKET_FROM_MAC)
Interference
(NOT_LAST_PACKET_FROM_RX)/wlan_phy_packet_from_rx()
(CCA_IDLE_FROM_RX)

Figure 5.3 Process model of 802.11n PHY in OPNET.

5.3.1 Capacity Calculation

Capacity calculation is performed in the external module **BPL interface** according to the formula (2.9). Moreover, the functions determine the number of optimal antenna pairs and antenna selection. Assume that the system is 2x2 MIMO with the channel matix for each subcarrier, as $\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}$.

For such a configuration, the channel capacities are calculated using (2.9) for

a) SISO channel with h_{11}

b) SISO channel with h_{22}

c) 2x2 MIMO channel with **H**

The capacities of the three configurations are compared, and the configuration offering the highest channel capacity is selected. If SISO is selected, 1 antenna will be used, otherwise the number of antennas is 2. If (c) offers the highest capacity, 2x2 MIMO will be selected for transmission between that transmitter-receiver pair.

5.3.2 Rate Adaptation

Rate adaptation is also performed in the external module **BPL interface** . The rate adaptation is done according to the channel states. When the station to be transmitted is determined, the rate is selected and parameters such as modulation, coding and number of antennas are determined. The data rate can be calculated as:

$$R(bps) = R_{raw} * R_c \tag{5.1}$$

with

$$R_{raw} = log_2 M * (1/T_S) * N_{sc} * N_a \tag{5.2}$$

where R_{raw} is the raw bit rate, R_c is the coding gain, M is the modulation level, T_S is the OFDM symbol duration, N_{sc} is the number of data subcarriers and N_a is the number of antennas. Accordingly, the following rates are achievable under the specified conditions in Table 5.2:

Table 5.2 Data rates and corresponding transmission modes

Modulation	Coding rate	Antenna #	Rate(Mbps)
BPSK	½	1	12
QPSK	½	1	24
BPSK	½	2	24
QPSK	¾	1	36
16 QAM	½	1	48
QPSK	½	2	48
16 QAM	¾	1	72
QPSK	¾	2	72
64 QAM	2/3	1	96
16 QAM	½	2	96
64 QAM	¾	1	108
16 QAM	¾	2	144
64 QAM	2/3	2	192
64 QAM	¾	2	216

Note that N_{sc} is 96 and T_S is 4 μsec throughout the simulations [1].

As explained in Section 2.2.1.2, for MIMO-OFDM systems the channel matrix is defined as N_C N_R x N_CN_T [19]. Rate adaptation is determined by comparing individual subcarrier capacities in (2.9). First, the transmission mode is determined according to the outcomes of capacity calculation explained in subsection 5.3.1. Having determined the transmission mode, SISO or MIMO, we start from the highest data rate for the defined number of antennas. Individual subcarrier capacities are compared with thresholds depending on the data rate. These thresholds are simply the total data rate divided by the number of subcarriers. If the number of subcarriers exceeding the threshold minus the number of subcarriers under the threshold is greater than a proportion (25 %) of the total data subcarriers, the rate is selected. Otherwise, the threshold comparison is repeated for the next highest data rate.

5.3.3 Channel Model

The properties of the wireless channel are modelled in the channel matrix ***H*** considering large-scale path loss and shadowing, and small scale multi-path fading effects. In this paper, we consider log distance path loss model and fading channel model "Channel B" defined by the Team Group n (TGn) [27] . The N_R x N_T channel matrix ***H*** duplicated in the extended MIMO-OFDM system as in (2.8) consists of a Rician component in addition to non-line-of-sight (NLOS) based coefficients.

In the wireless channel, the log-distance path loss is modelled with path loss exponent of 2 within a distance of 5 meters from the transmitter and 3.5 for distances larger than 5 m. Log-normal shadowing term is taken as 3 dB up to 5 m and 5 dB after 5 m. In other words, path loss, *PL* is calculated as

$$PL(d) = L_0 + 10\alpha \log(d / d_0) + \mathrm{N}(0, \sigma^2) \qquad (dB) \ , \tag{5.3}$$

with

$$\alpha = \begin{cases} 2 & ,d < 5m \\ 3.5 & ,d > 5m \end{cases} \qquad \sigma = \begin{cases} 3 & ,d < 5m \\ 5 & ,d > 5m \end{cases} , \tag{5.4}$$

In this channel model, the fading characteristics between individual antenna pairs are spatially correlated Therefore, NLOS terms of the channel matrices are formed by the multiplication of a matrix with independent identically distributed (iid) complex Gaussian random variable elements with transmit and receive correlation matrices. The correlation matrices depend on the angular spread. The 802.11n general channel model is assumed as the superposition of two clusters. For each cluster, the channel matrix can be composed as a line-of-sight (LOS) component and a non-line-of sight (NLOS) component.

$$\mathbf{H} = \sqrt{\frac{K}{K+1}}\mathbf{H}_{\mathbf{LOS}} + \sqrt{\frac{1}{K+1}}\mathbf{H}_{\mathbf{NLOS}} \tag{5.4}$$

K is taken as 0 for Channel B. The NLOS components X_{ij} between receiver antenna *i* and transmitter antenna *j* is dependent on the receive and transmit correlation matrices R_{rx} and R_{tx}:

$$H_{NLOS} = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}, \tag{5.5}$$

$$H_{NLOS} = (R_{rx})^{1/2} H_{iid} ((R_{tx})^{1/2})^T, \tag{5.6}$$

where H_{iid} is consists of iid zero-mean complex Gaussian variables. The elements of the correlation matrices depend on the power angular spread (PAS), where the PAS is the second moment of the angular spread (AS). The correlation coefficients depend on the separation between the antennas. Further details can be found in [27].

The power delay profile of the two clusters of the Channel model is summarized in Table 5.3:

Table 5.3 Power Delay Profile of Channel B

	Delay	0	10	20	30	40	50	60	70	80
Cluster I	Rel. Power(dB)	0.0	-5.4	-10.8	-16.2	-21.7				
Cluster II	Rel. Power(dB)			-3.2	-6.3	-9.4	-12.5	-15.6	-18.7	-21.8

The rms delay spread τ_{rms} is calculated by the formula [11]

$$\tau_{rms} = \sqrt{\frac{\sum_{k=1}^{N} \tau_k^2 \sigma_k^2}{\sum_{k=1}^{N} \sigma_k^2} - \left(\frac{\sum_{k=1}^{N} \tau_k \sigma_k^2}{\sum_{k=1}^{N} \sigma_k^2} \right)^2}, \tag{5.7}$$

where τ_k is the multipath delay and σ_k^2 is the corresponding power for that delay value. Inserting values from Table 5.3 into (5.7), we calculate the delay spread as 15 nsec. If we assume that coherence bandwith is $1/5\,\tau_{rms}$, the maximum data rate supported over the channel is 13.3 Mbps. For other channel models with higher delay spread, this value is further reduced. Hence, the usage of OFDM enables us to operate over channels with 20 MHz or 40 MHz(in our case) without experiencing frequency-selective fading, since the coherence bandwith is much greater than the bandwith of one OFDM carrier, which is 3.125 KHz. Due to low speeds of WLAN users, slow fading is assumed. The Doppler frequency is 5 Hz, leading to a coherence time of 100 msec from the $1/2\,f_d$ formula [18] .

In the next subsection, we evaluate the performance of the above algorithms, also considering different data rates defined by the IEEE 802.11n and aggregation transmission.

CHAPTER 6

PERFORMANCE ANALYSIS AND SIMULATION RESULTS

The results of extensive simulations performed in this work for the evaluation of scheduler performances are given in this chapter. Comparisons were carried out in sections 6.1 to 6.4 for the following algortihms:

- Predictive Scheduling with Time-Domain Waterfilling (P-WF)
- Predictive Scheduling with Access Guarantees (P-AG)
- Aggregate Opportunistic Scheduling (AOS)
- Aggregate Discrete Opportunistic Scheduling (ADOS)
- Capacity Queue Scheduling (CQS)
- Proportional Aggregate Opportunistic Scheduling (P-AOS)
- Opportunistic Autorate (OAR)
- Maximum Rate Scheduling (MRS)
- Shortest Remaining Processing Time First(SRPT)
- Proportional Fair Queueing (PFQ)
- Longest Queuing (LQ)

Throughout simulations, either parameters such as load and transmission opportunity limit have been varied or performance under different topologies for identical parameters have been compared. The effects of relaying on the AOS, CQS and LQ algorithms are evaluated in section 6.5.

6.1 Throughput Analysis

Downlink traffic is modeled by fixed size (1024 bytes) packets that arrive due to the Poisson distribution. Similar load levels for all stations are assumed, and the AP buffers are assumed large enough. In our simulations, topologies with an AP and 12 stations in an area which is approximately a circle with a radius of 25 m from the AP are used with Topology #1 as shown in Figure 6.1 .

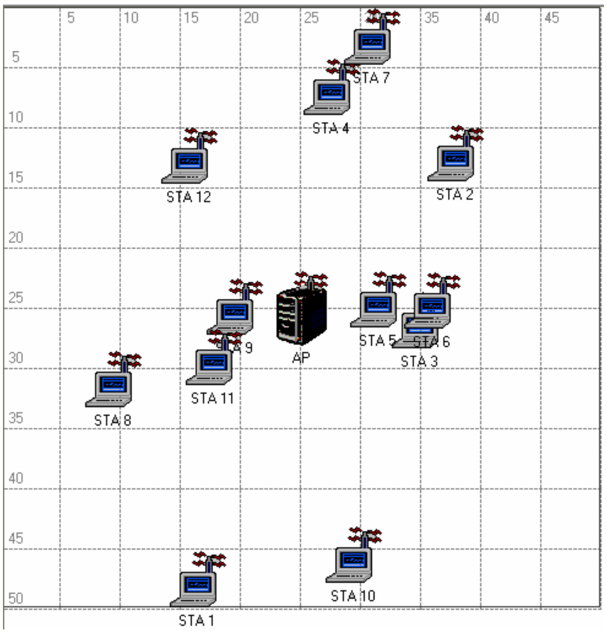


Figure 6.1 Topology #1.

We first evaluate the scheduling algorithms under varying load, changing the network load between 50 Mbps and 200 Mbps. Our results in terms of network throughput for Topology #1 are demonstrated in Figures 6.2 and 6.3. For all scheduling disciplines, the throughput and load is about the same as long as the arrival rate is below the network service rate. The “network service rate” depends on two factors: the physical medium, i.e., the maximum data rates that can be supported, and the MAC efficiency. Network throughput is measured as the total number of bits delivered to all of the users per unit time.

The effect of aggregation on throughput considering existing scheduling algorithms from literature, namely, MRS, PFQ and LQ is shown in Figure 6.2. Frame aggregation significantly increases throughput and the amount of enhancement depends on the scheduling algorithm in concern. For instance, the LQ algorithm outperforms the MRS algorithm using frame aggregation, due to the fact that with the MRS algorithm, users with higher channel capacities are served frequently and their queues do not fill up. As a result, the aggregate sizes of transmitting to these users are low, and low aggregate sizes yield reduced throughput. On the other hand, the service rate is shared equally among the users; consequently, the per user service rate is much lower, leading to larger aggregate sizes for the LQ algorithm. Transmitting with larger aggregate sizes yield higher throughput prevailing over using higher data rates. The capacities of the stations with poor channel conditions showed a higher deviation around their mean values as compared to channel capacities of the better conditioned stations for this topology and our channel model. This lead to the preference of users associated with poor-conditioned channels more frequently with the PFQ algorithm, leading to lower throughput performance as compared to the LQ algorithm. When the arrival rate exceeds the service rate, the total throughputs of the scheduling algorithms deviate from each other. It is observed that the throughputs of the LQ and PFQ algorithms are saturated after some load value, whereas the throughput of the MRS algorithm continues to increase with load. However, total load should increase significantly to result in better throughput performance. This outcome is mainly due to the fact that the MRS algorithm selects best users with highest data rates but transmitting with larger aggregate sizes as load is increased. In short, frame aggregation increases throughput of known schedulers, by a factor of 2.5-3.5.

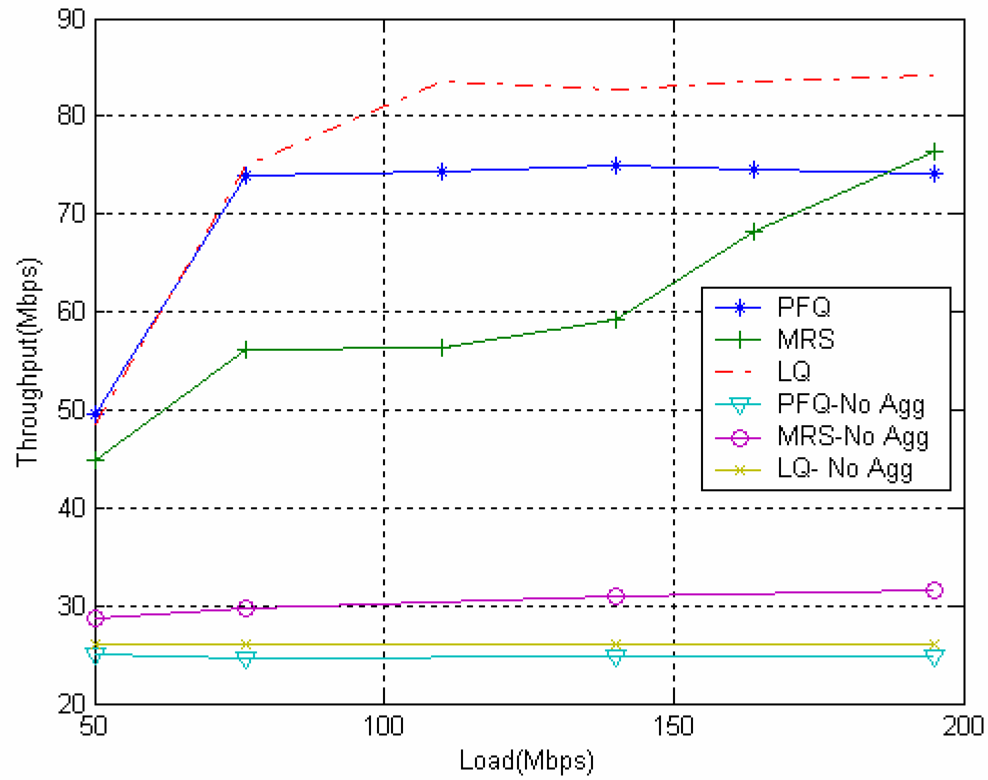


Figure 6.2 Performance of existing schedulers with and without frame aggregation.

Next, we compare our proposed schedulers AOS, ADOS, CQS, P-WF, P-AG and P-AOS with existing scheduling algorithms, LQ, MRS, PFQ, SRPT and OAR, all employing frame aggregation. It can be readily inferred from Figure 6.3 that out of the queue aware algorithms AOS and ADOS algorithms significantly outperform the existing algorithms, namely, up to 21 % over LQ, 35 % over MRS/PFQ and 53 % over SRPT. The SRPT algorithm has the smallest throughput with aggregation, although the stations selected for the SRPT are similar to the stations selected in MRS. SRPT favors the best positioned stations with lowest queue sizes. This leads to smaller aggregate sizes, resulting in lower throughput. Note that there are two versions of OAR in the Figure 6.3. The definition of the OAR algorithm states that the aggregate size is defined as the ratio of the data rate of the station over basic rate. Since basic rate is 24 Mbps in our simulations, this leads to small aggregate sizes and very low throughput. In order to improve throughput of OAR, we also applied the algorithm with a basic rate of 12 Mbps, which is the lowest defined basic rate. If the basic rate is further decreased to very low values, the throughput of the OAR algorithm starts to decrease again, since the temporal access share of the users with poor channel conditions increases.

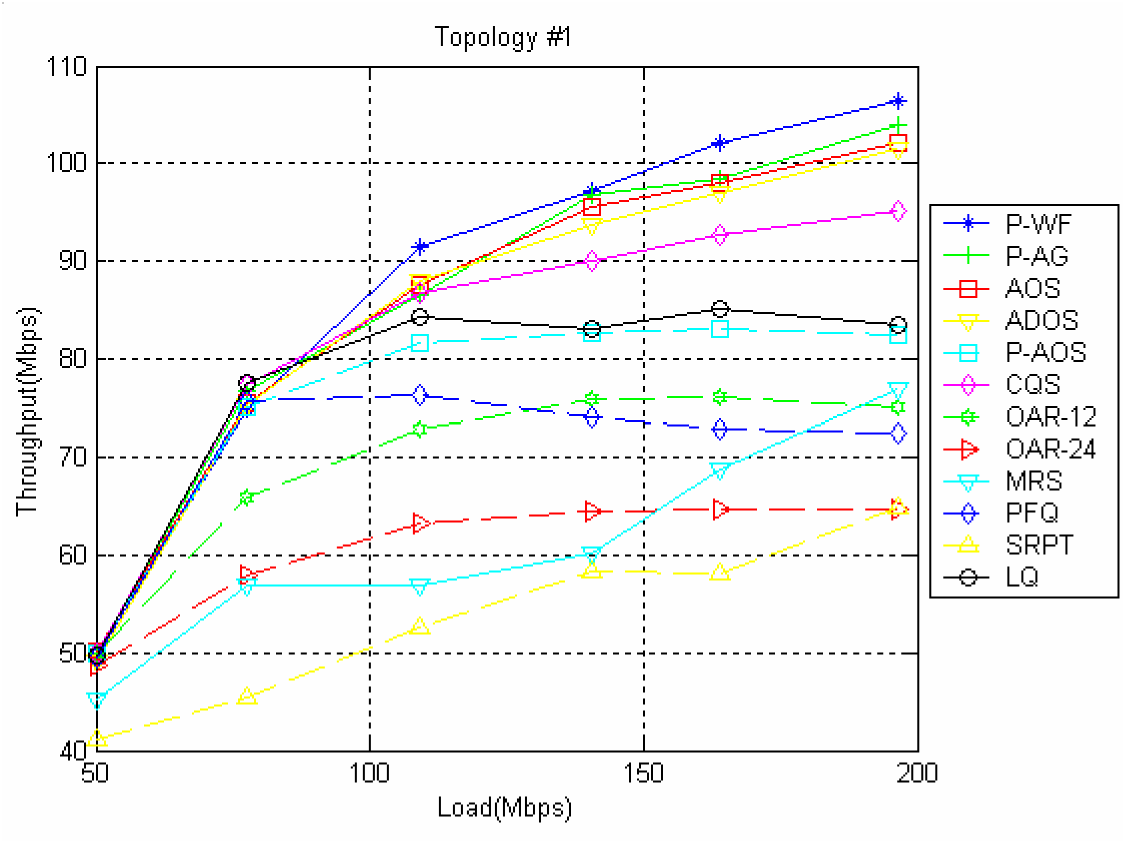


Figure 6.3 Performance of proposed and existing schedulers with frame aggregation.

Out of the queue aware algorithms, the AOS and ADOS algorithms have the advantage of possessing the most explicit insight about the system performance, since they use the best approximation to the throughput, considering the effects of both the physical medium and MAC efficiency. CQS shows 10% poorer performance as compared to AOS and ADOS, but still outperforms the LQ, MRS, PFQ, OAR and SRPT algorithms. The algorithm favors good-conditioned channels in the general sense; however, those users with high quality links are not continuously served, causing their queues to grow, leading up to large aggregate sizes. Our schedulers based on queuing theory, namely P-AG and P-WF outperform queue aware schedulers. P-WF offers the highest throughput out of all of the algorithms, achieving a further 4 % improvement in throughput over AOS.

The throughput of the P-AOS algorithm exceeds the throughput of the PFQ algorithm by about 10 %. If all users were saturated and the aggregate sizes were large, the PAOS algorithm would have converged to the PFQ algorithm, since the average throughput would be close to the average capacity and the instantaneous throughput would have been close to the instantaneous capacity. However, when instantaneous throughput is compared to its average value, users with very high capacity channels are selected more frequently since throughput is more sensitive to variation sizes in aggregate size when the aggregate sizes are small. In initial stages of simulations, when the aggregate sizes are low since the queues are not filled yet, the advantage of the sensitivity for very high capacity users prevails over the fact that the channel capacities of poor channel capacities vary more over their average values. Since high capacity users are selected frequently, their queues do not pop up, leading to operation with low aggregate sizes and low average throughput values. Since average throughput values remain low, the metrics of high channel capacities increase to high quantities more frequently than other channels, which enable high channel capacity users to be selected more. Even though users with highest channel capacities are favoured, remaining high channel users are not selected frequently. Furthermore, the high capacity users are transmitted with low aggregate sizes, leading to reduction in throughput.

In order to better understand the behaviours of schedulers, we observe the variation of physical data rates and aggregate sizes with respect to changing load levels in the simulations. In Figure 6.4, we present the time-averaged physical data rates used by the schedulers are plotted. By plotting the average data rate graphic, our aim is to show what would be the throughput if no MAC and PHY overhead were present.

Note that we present time-averaged data rates instead of simply calcuating the ensemle average of used data rates. The reason for doing so is that time averged data rates provide better insight about throughput, which is the amount of data transmitted successfully throughout the network per unit time. Calculating arithmetic data rates is misleading since transmission of frames with low data rates takes longer duration compared with frame transmision using higher data rates. For example, assume that we transmit two equal-size frames with data rates 100 Mbps and 50 Mbps, respectively. The arithmetic average is 75 Mbps. However, when we calculate the number of bits transmitted over unit time, the time averaged data rate is given by

$$TADR = \frac{2}{\left(\frac{1}{100\times10^{6}} + \frac{1}{50\times10^{6}}\right)} = 66.\bar{6}\times10^{6}\,bps\,, \qquad (6.1)$$

which is lower than the arithmetic data rate. The fact that time averaged data rate is lower than arithmetic avergae data rates is also valid for any simulation set. If no overhead was present, the time avereged data rates would have been equal to throughput.

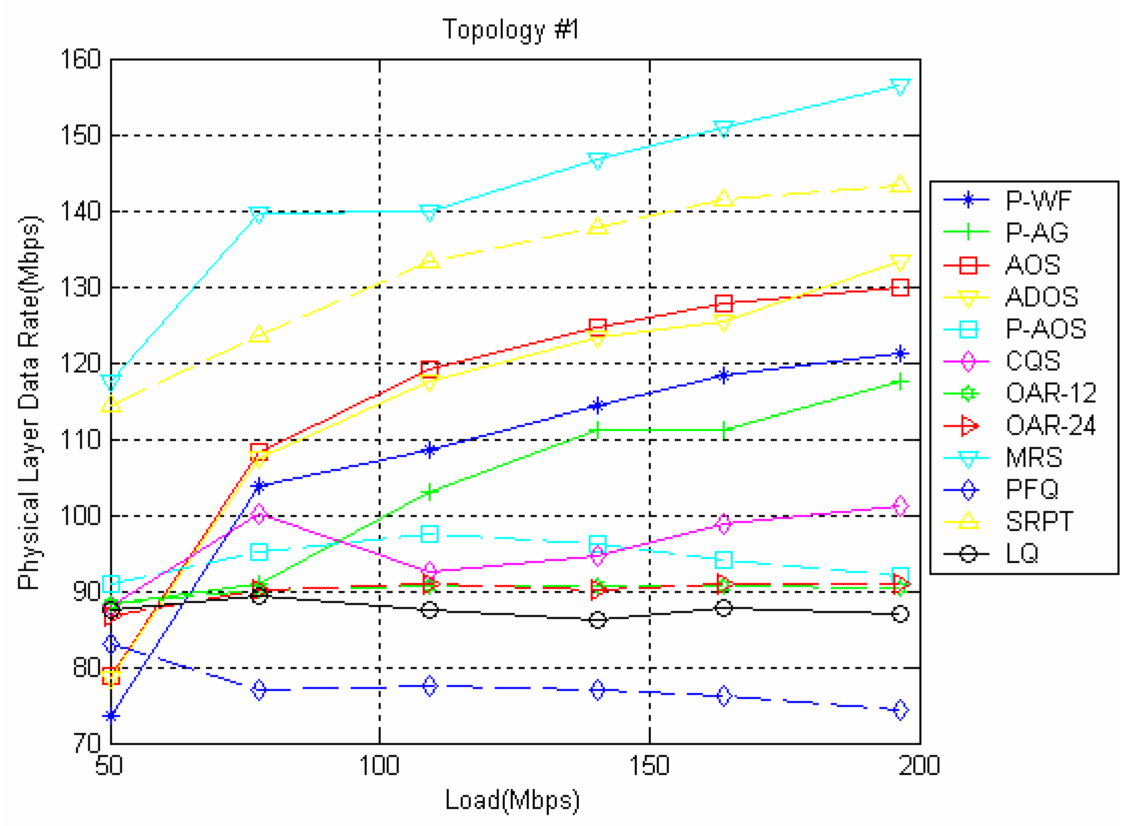


Figure 6.4 Average data rate depending on load.

As expected, the LQ algorithm operates over lower data rates compared to opportunistic schedulers and the most “opportunsitic” schedulers such as MRS and SRPT operate with the highest data rates, since they prefer users with higher capacity. PFQ uses data rates even lower than LQ due to the characteristics of the wireless channel which varies more over its mean behavior at higher distances. AOS, ADOS, P-WF and P-AG operate with lower data rates compared with MRS since queue size is also important in addition to channel capcaity, with P-AG lower than AOS,ADOS and P-WF since service is guranteed even for users with very poor channel conditions. The data rates of the CQS algorithm is between the data rate of these four proposed schedulers and LQ, which doesnt take capacity into account in the scheduling decision.

Even though the MRS and SRPT algorithms use the highest data rates, they cannot offer the highest throughput. This is due to the fact that since the aggregate sizes are very low compared to other schedulers, as shown in Figure 6.5, which depicts average aggregate sizes of users with different schedulers. MRS is slightly higher than SRPT since SRPT also favors short queues from (2.13) . As expected, for saturated scenarios the aggregate sizes of the CQS and LQ algorithms are the highest, since the queues of all users tend to be full. The aggregate size of the CQS algorithm is slightly higher than LQ . The reason is due to the fact that the transmission opportunity is limited by 10 msec. For this transmission duration limit, all of the packets intended for aggregation may not be transmitted for users with low data rates since the transmission durations are longer. The LQ algorithm selects low capacity users with a higher frequency compared with CQS, resulting in a slightly lower average aggregate size. The aggregate sizes of the OAR algorithms are low, with higher aggregate size with basic rate of 12 Mbps. These low aggregate sizes cause throughput to be very low even though the data rates are higher than LQ. Aggregate sizes of the P-AOS is lower compared with the average aggregate size of the PFQ algorithm since users with very high capacity are served with low aggregate sizes.

Although the time averaged data rate of P-WF and P-AG is lower than AOS and ADOS, throughput of the predicitve schedulers exceed the throughput of AOS and ADOS since they use larger aggregate sizes. The effect of aggregate size on throughput is more important for transmissions with high data rate since the transmission duration of a single packet is shorter as compared to low data rates. Since the packet transmission durations are shorter, in order to achieve the same level of MAC efficiency, wihch was defined in Chapter 2 as the ratio of throughput to actual data rate, larger number of aggregated packets are required. Thus, the fact that the average aggregate sizes are lower for AOS proves a penalty in terms of throughput compared with the P-WF and P-AG algorithms.

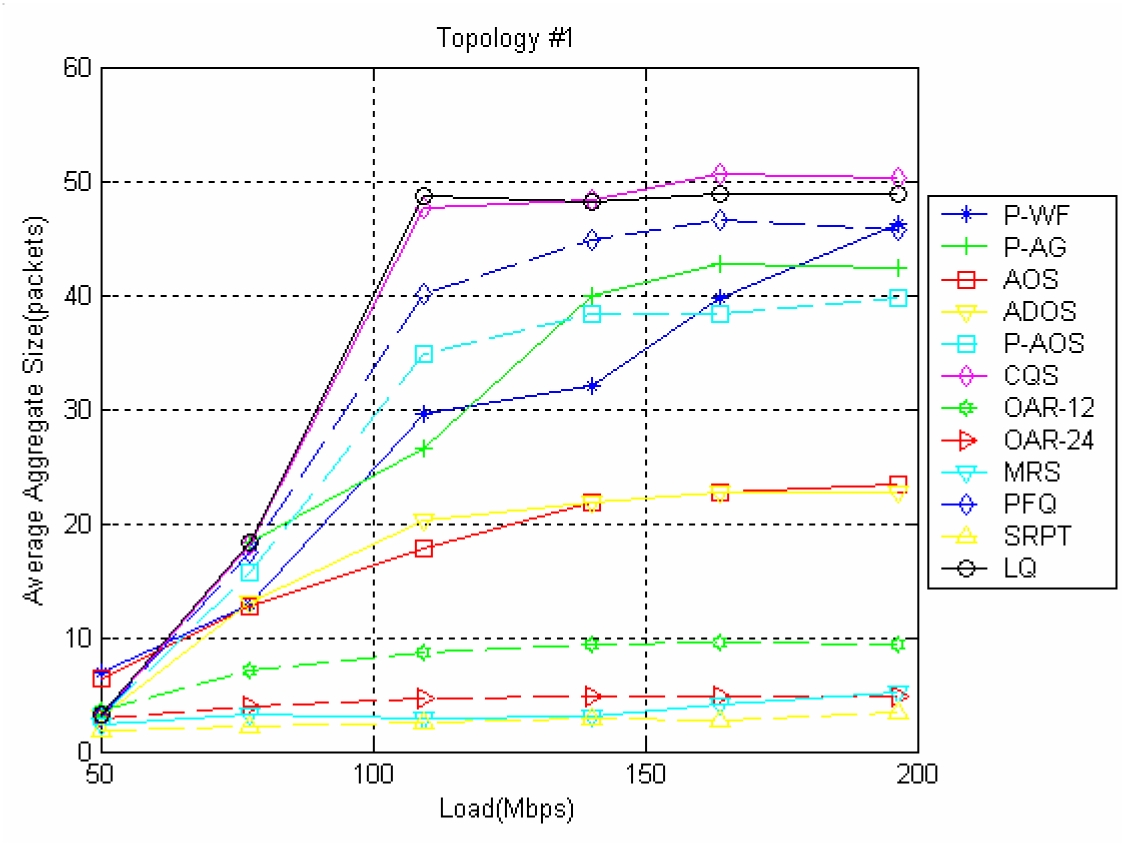


Figure 6.5 Average aggregate size depending on load.

Figure 6.6 depicts the MAC efficency of each scheduler, where the actual throughput and time averged data rates are plotted togerther for the maximum load level. LQ and CQ operate with highest eficiencies due to large aggregate sizes, however throughput is low since selected physical layer data rates are low. SRPT and MRS are the most inefficient schemes. Despite operating at highest physical layer data rates, the achived throughput levels are half or less than half of available capacity.

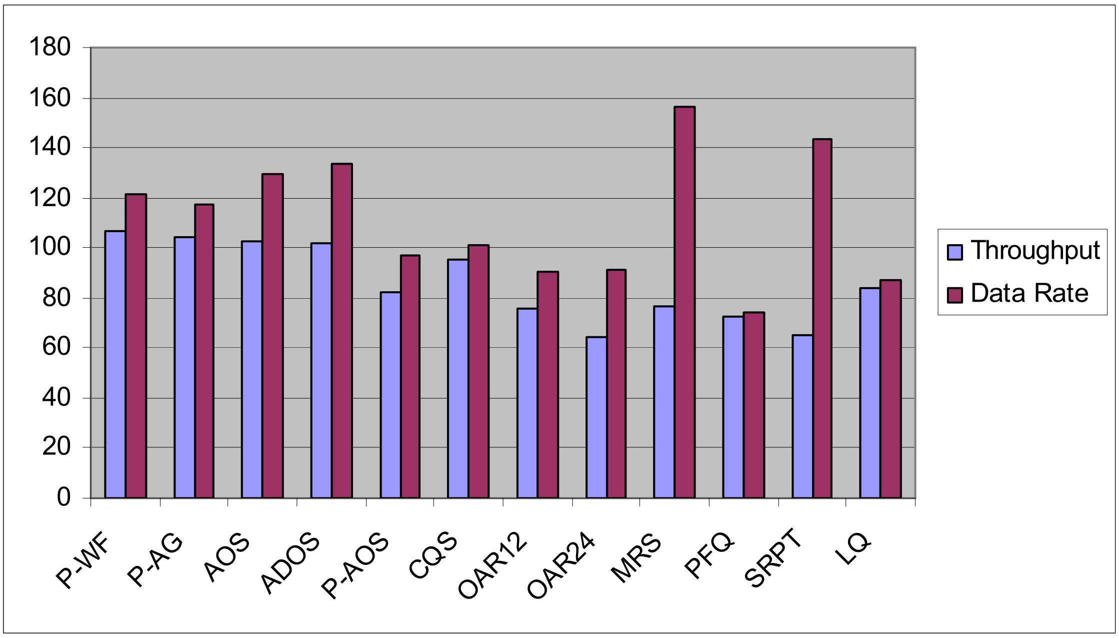
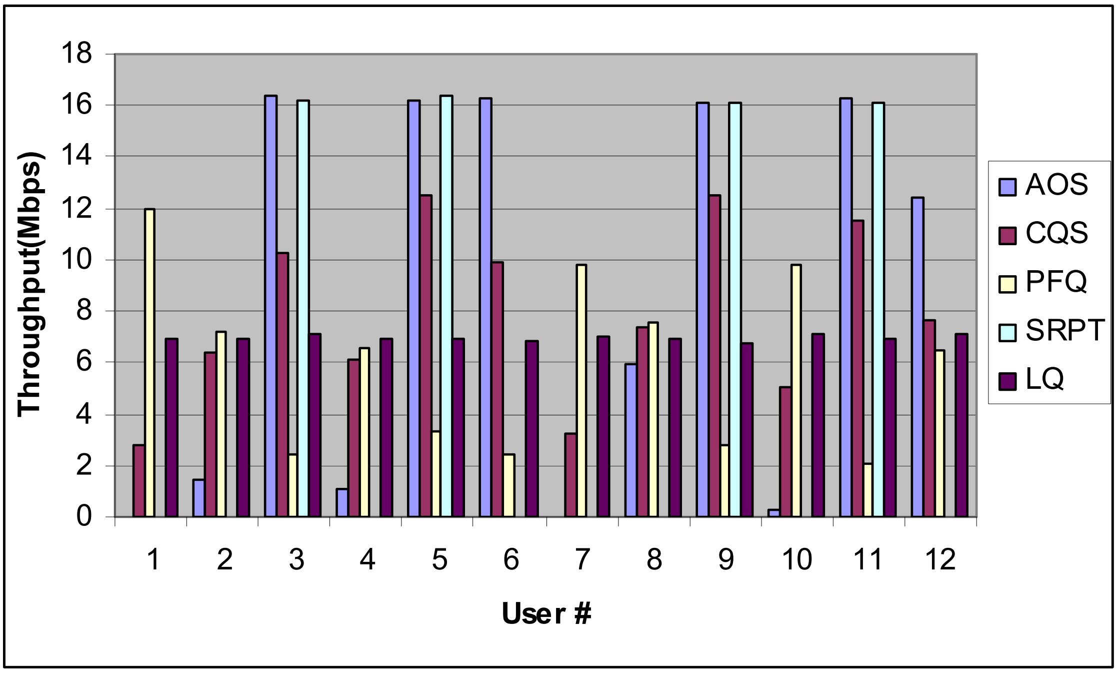
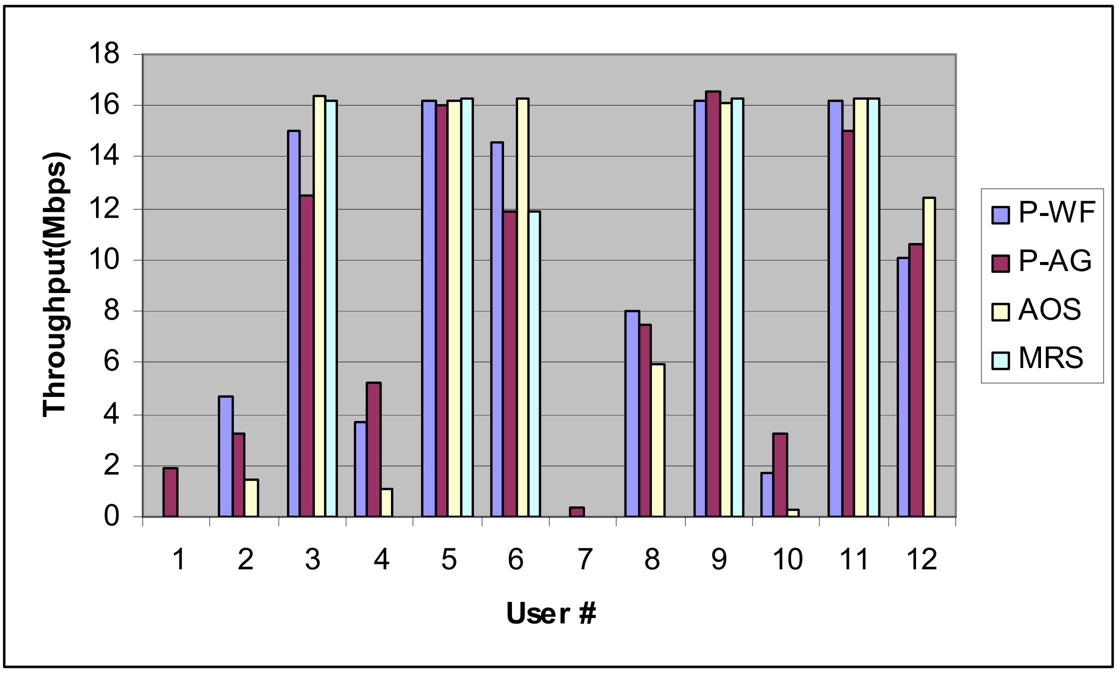


Figure 6.6 Throughput vs. Data Rate.

After presenting results which are concerned with overall network operation, we present the throughput performances of individual users at the maximum load level, 200 Mbps. For better readibility, results were stacked into two graphic pairs and some algorithms were not presented. Figure 6.7a and 6.7b present the throughputs. MRS and SRPT do not serve many users. AOS and P-WF may also cause some users to starve. P-AG provides access for every user, leading to a non-zero throughput for each user. LQ leads to equal throughput. The users PFQ favors are the users which are not favored at other opportunistic algorithms.



a)



b)

Figure 6.7 Individual User Throughputs for Topology #1.

In Fig. 6.8, we study the performance of all scheduling disciplines as a function of the allowed maximum aggregate size. For this, we varied the maximum transmission opportunity (TXOP) duration, which determines the maximum aggregate size. Simulations were carried for an aggregate load of 200 Mbps. For low TXOP values, all opportunistic schemes perform better than LQ, since LQ cannot take advantage of large aggregate sizes. As the TXOP is increased, larger aggregate sizes are allowed, LQ becomes more efficient than MRS, PFQ and SRPT. Our schemes P-WF, P-AG, AOS, ADOS and CQS outperform all other schedulers for all TXOP values. Note that the throughputs of algorithms saturate after some TXOP duration, which limits the performance by aggregation.

Average aggregate sizes of the algorithms depending on the maximum TXOP limit are presented in Figure 6.9. For very low TXOP values, algorithms which transmit to users with poor channel conditions frequently such as LQ fail further. Since transmission with lower data rates results in longer packet transmission durations, the aggregate sizes used are very low. Note that the aggregate sizes are slightly lower than 63 for the high TXOP limit. This is attributed to channel errors where some packets are retransmitted. Even though the temporal effect of these retransmissions may be negligible, they reduce the calculated aggregate size.

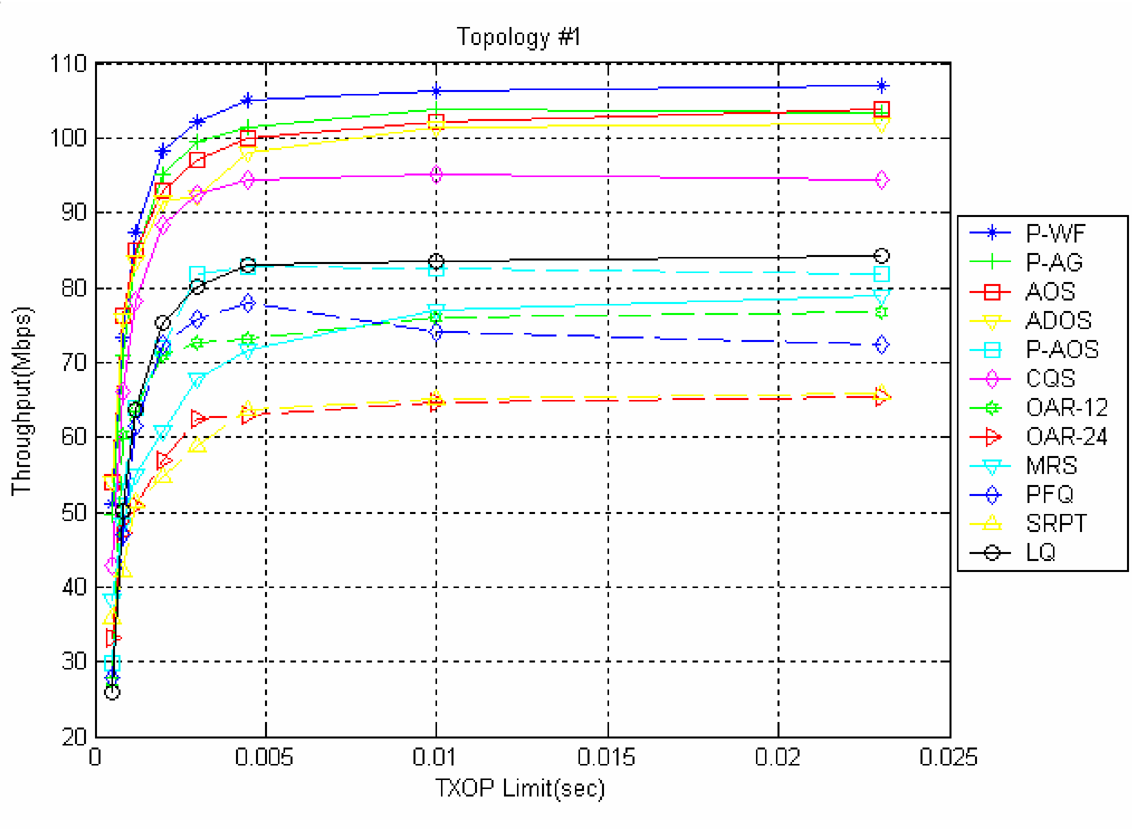


Figure 6.8 Performance of schedulers as the TXOP Limit.

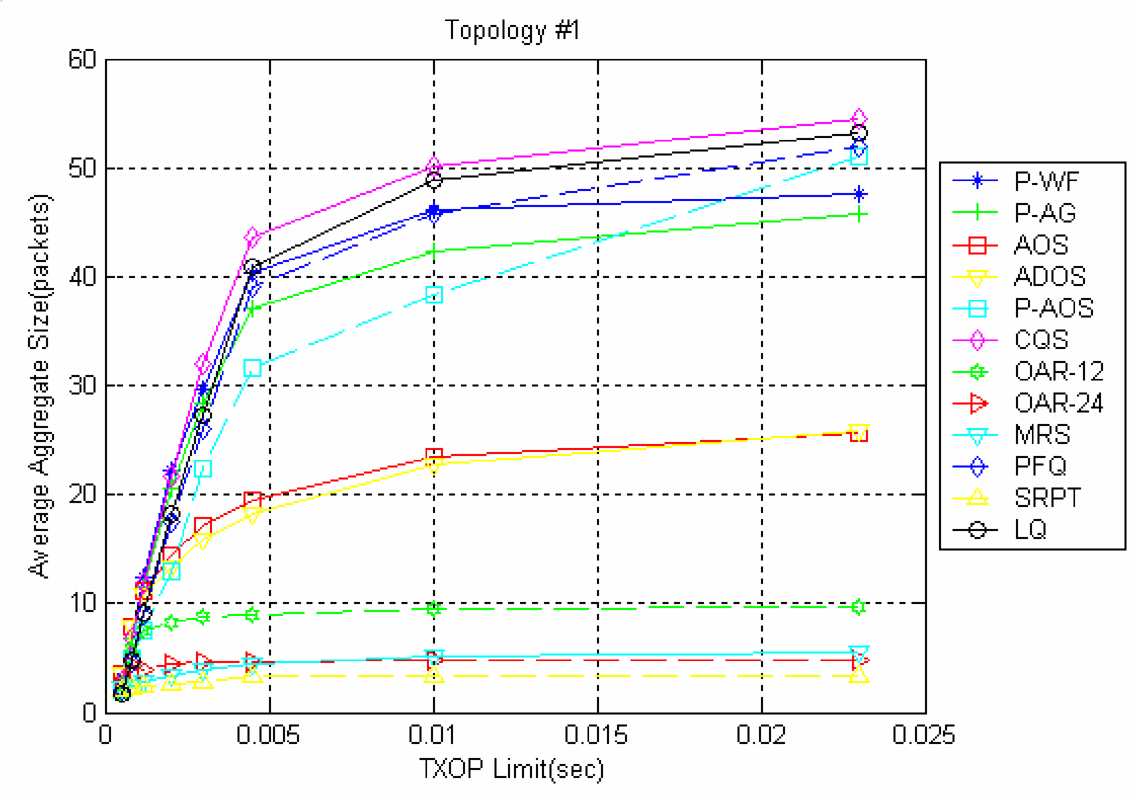


Figure 6.9 Aggregate size vs. TXOP Limit.

6.2 Delay Analysis

In this section, we present delay performance of the schedulers. The delay results presented in Figure 6.10 are the sample mean of average delays experienced by each user. The simulations were carried out for the duration of 5 seconds. When we consider the scheduling metrics in more detail we observe that the nature of the LQ algorithm tends to prevent users to starve a lot from transmission, which means the algorithm intends to prevent very large delays. For this reason, the delays of individual users are almost identical for the LQ algorithm. On the other hand, opportunistic schedulers cause user delays to differ from eachother since usually high capacity users are favored. The delays of such users are low, while low capacity users encounter very large delay. Especially in pure opportunistic schedulers such as MRS and SRPT, some users may never be selected if they do not maximize the scheduling metric. While evaluating delay performance for those users, we assume that their average delays of those users are equal to the simulation duration, i.e. 5 sec. When we observe the overall effect, for Topology #1, the P-AG algorithm provides the lowest mean user delay. Since LQ and CQ prevent users to starve too much, their delay performance is better than AOS and P-WF, which may cause some users to starve. MRS and SRPT perform poorly since lots of users suffer. On the other hand, the P-AG algorithm, which guarantees service for every user, provides the best delay performance over all algorithms. Even if the delay for poor users may be high, the fact that they are bounded results in a low overall delay. Throughput and delay are plotted together in Figure 6.11. The P-AG algortihm offers the best performance when delay is taken into account as well as throughput.

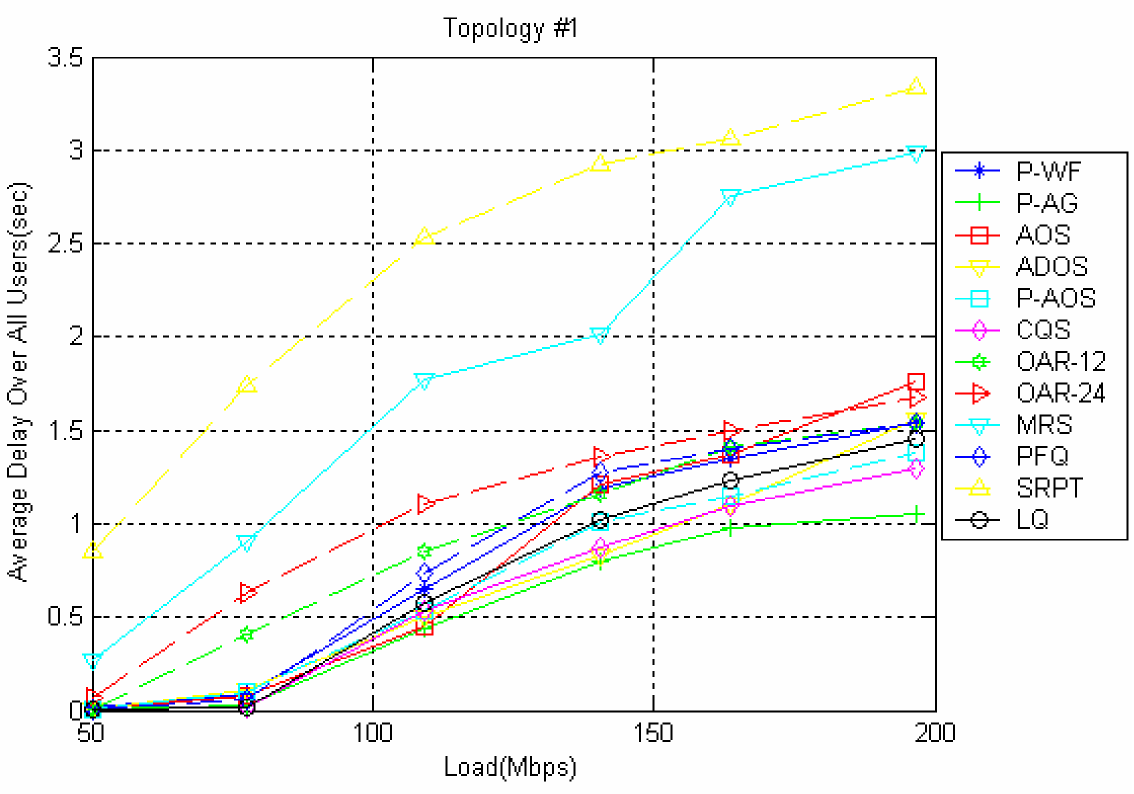


Figure 6.10 Mean user delay vs load.

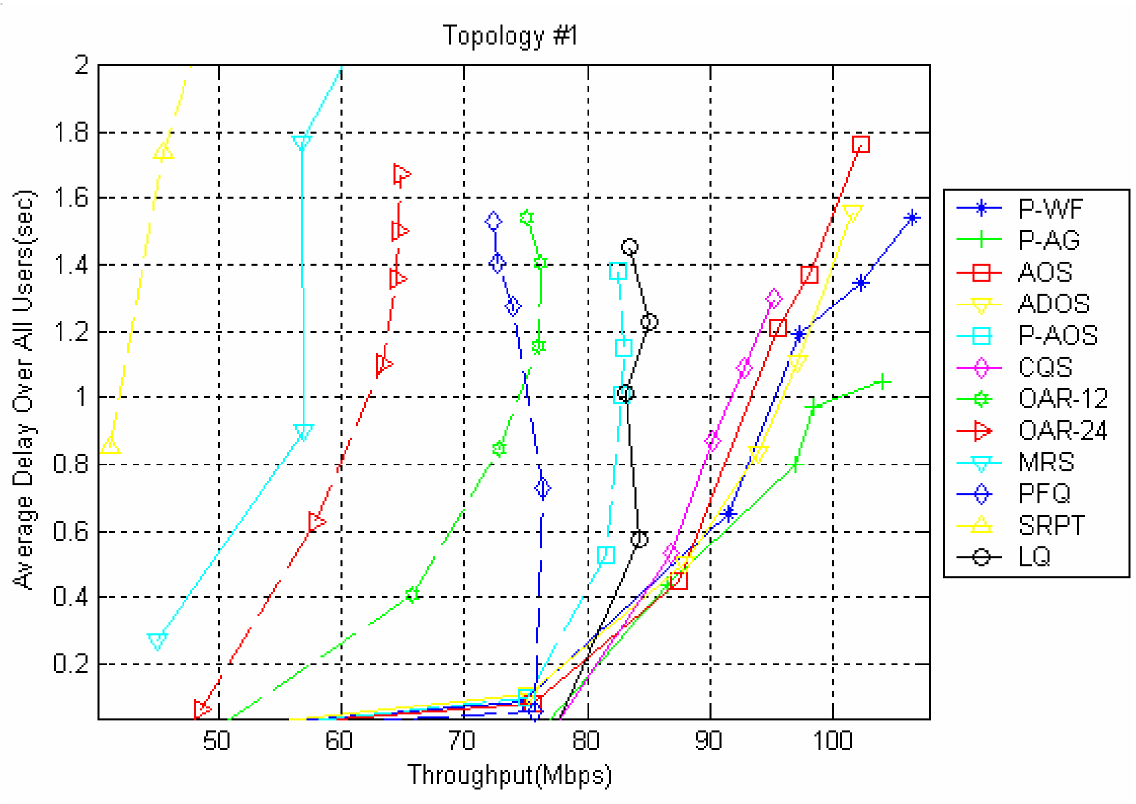
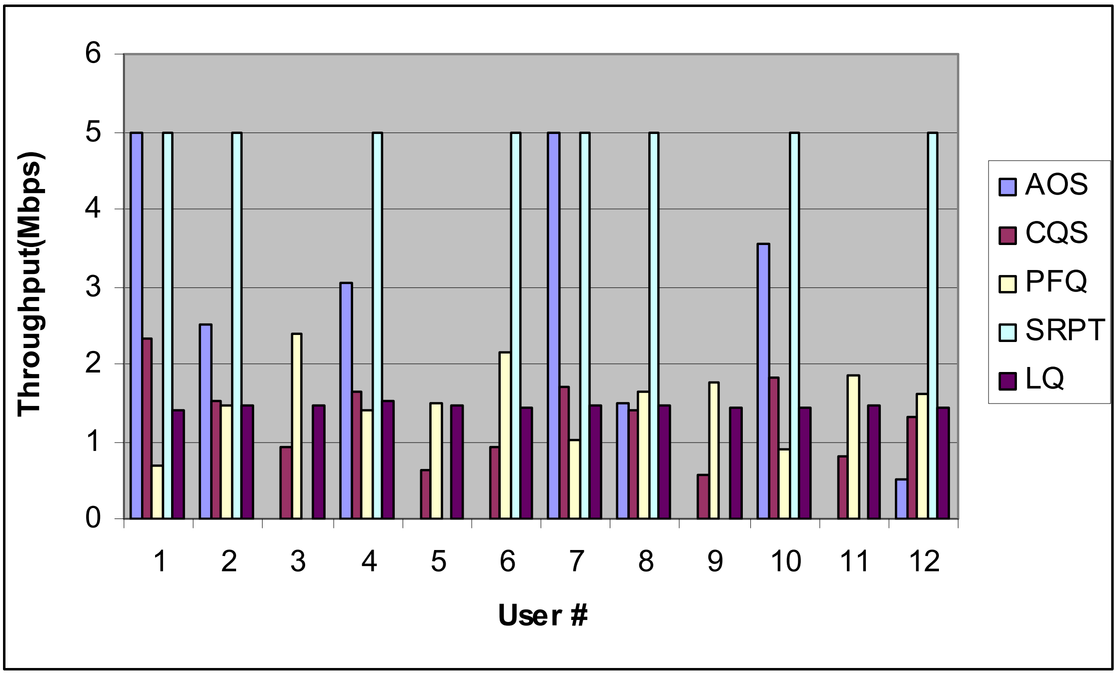
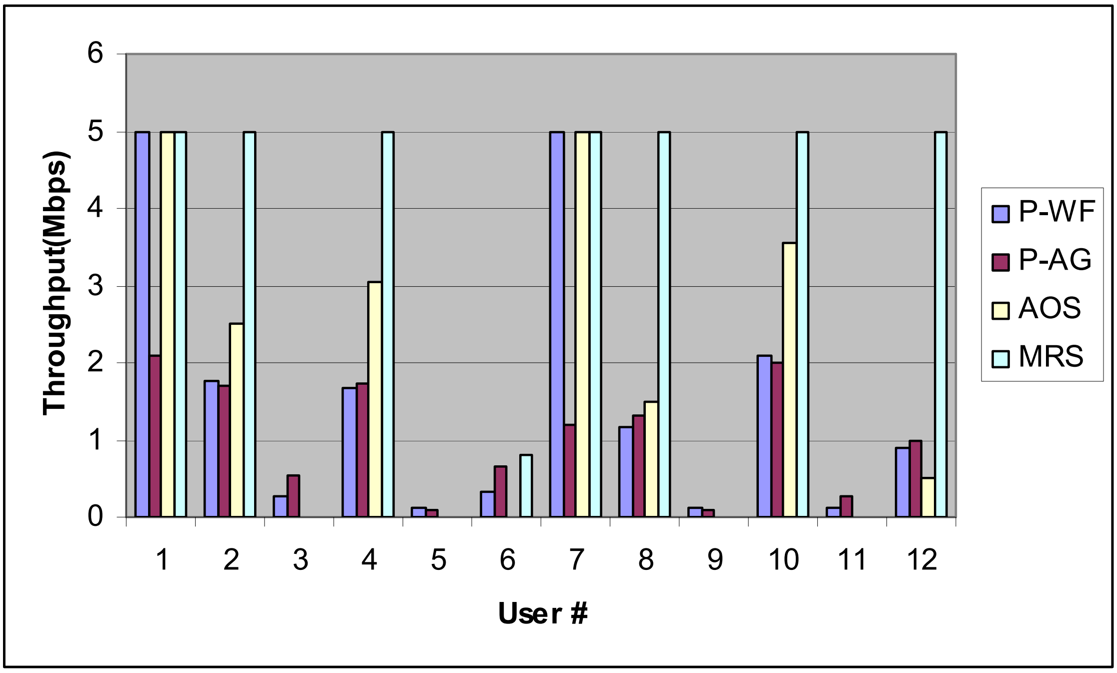


Figure 6.11 Delay vs. Throughput.

Figures 6.12a and 6.12b show the average delays of individual users. Even though good channel users are almost instantly served for MRS,SRPT and AOS, users with lower channel capacity experience large delays. In LQ and CQ, all users have bounded delay but the good users experience more delay. The best compromise occurs for P-AG, which provides the best overall average delay.



a)



b)

Figure 6.12 Individual Average User Delays for Topology #1.

6.3 Fairness Analysis

Having discussed the performance of algorithms in terms of average overall throughput and delay, next we present the fairness performance of the scheduling algorithms for the given topology. To evaluate fairness, we first compute the average station throughput, S_{av}. Next, we compute the standard deviation, *σ,* of the set of throughputs of individual users. That is,

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\overline{S_i}^2 - (S_{av})^2, \qquad (6.2)$$

where $\overline{S_i}$ corresponds to the average throughput of user *i* and *N* is the number of stations in the system. (This formula is valid under the assumption that every station has identical traffic characteristics). We define the unfairness measure as the ratio of the standard deviation of station throughputs to the mean value of station throughput as follows:

$$UF = \sigma / S_{av} \qquad (6.3)$$

It is obvious that the larger *UF* gets, the distribution of throughputs among stations becomes more unfair. Using the definition of this unfairness measure, a picture of the fairness performances of the algorithms under varying load can be seen in Figure 6.13.

Our simulations indicate that, the LQ algorithm performs best in terms of fairness. SRPT and MRS algorithms are poorest, as expected, since they favour users with users with high capacities strictly. In addition to the superior throughput performance, our algorithm AOS yields a more balanced throughput distribution among individual users as compared to MRS and SRPT. The ADOS algorithm offers slightly more fair distribution than AOS. This is due to the fact that ADOS quantizes the data rates to certain levels, which results in increased emphasis on queue sizes and enhances fairness. CQS algorithm shows best fairness among our proposed schemes, closest to LQ, due to the trade off between channel capacity and queue size in its metric. The fairness performance of the OAR algorithm is similar to the CQ algorithm, since the ratio of total packets served over all transmitted packets is similar for both algorithms, roughly proportional to their capacities. As for our predictive schedulers, we observe that P-WF yields a more fair allocation than AOS but P-AG significantly outperforms AOS. This is due to the fact that every user is guaranteed service, even if the actual access proportion for users with poor channels may actually be very low.

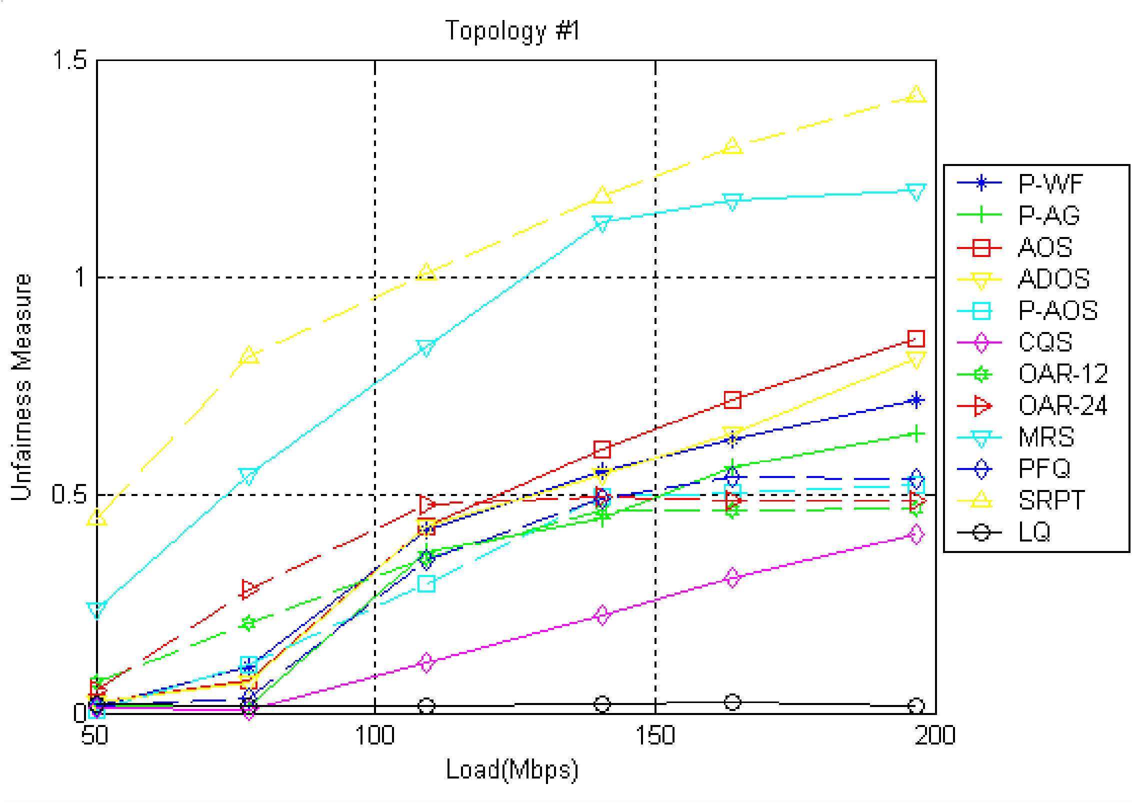


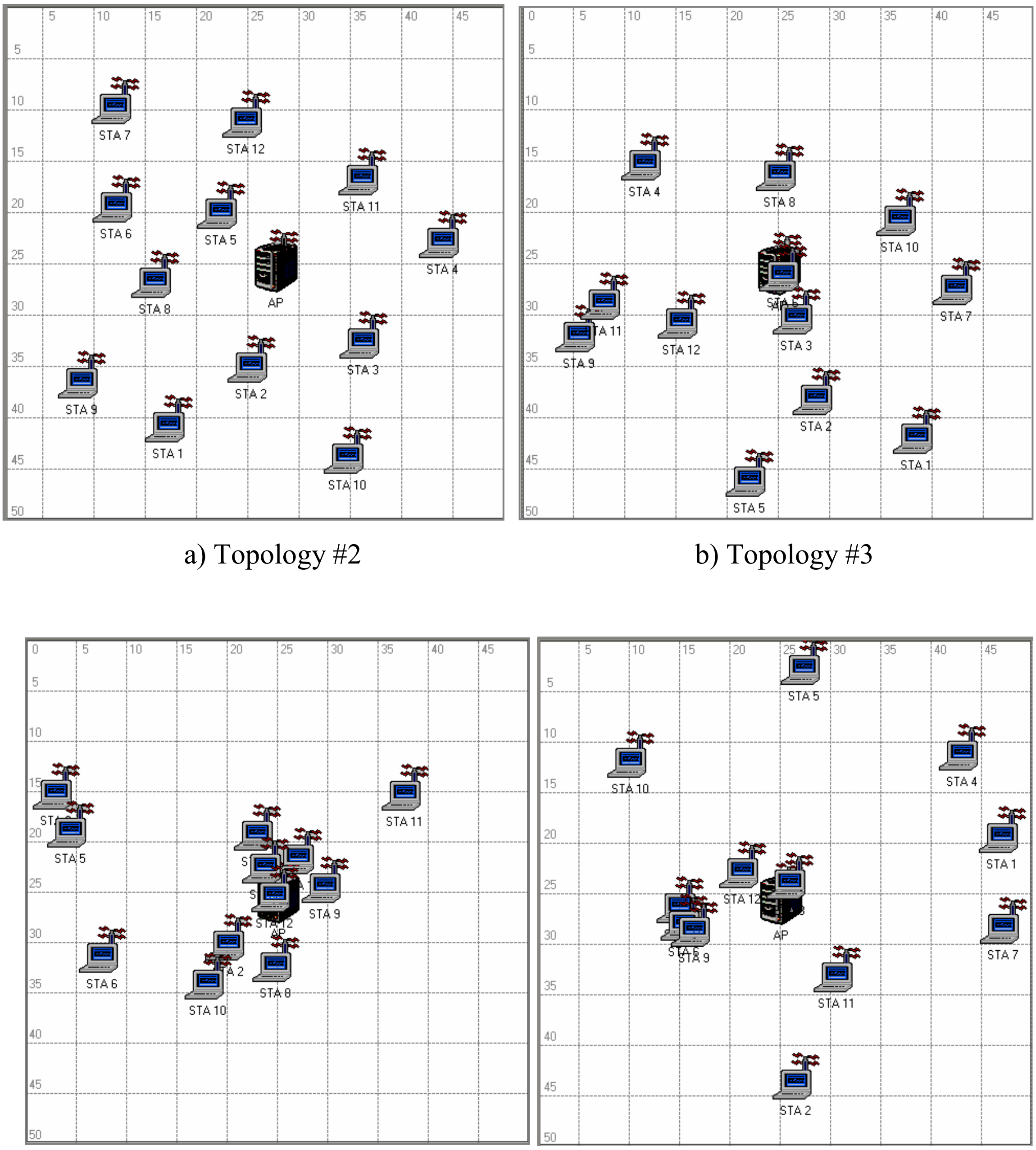
Figure 6.13: Fairness performance under varying load.

A fact inferred from Figure 6.13 is that throughput distribution among users become more unfair as load is increased. When load levels are low and all users are served, there is no issue of fairness problems since all of the traffic requirements are satisfied. As load is increased, the network cannot fully support all of the input traffic and users with high channels are selected more. When load is very high, the more "opportunistic" algorithms which give high priority to high capacities are able to make use of the high capacity users since the packets for those users are emptied after serving them more, since the number of packets to serve increases with load. This causes unbalanced throughput distribution among users, leading to a higher unfairness measure.

6.4 Performance with Different Network Topologies

6.4.1 Performance at Full Load

Our third simulation set is related with the performances of the scheduling approaches with different topologies shown in Figure 6.14 a-d. in addition to Topology #1.



a) Topology #2

b) Topology #3

c) Topology #4

d) Topology #5

Figure 6.14 Topologies Used.

Note that out of the five topologies, in Topology #2 is the topology with a more uniform user location, without stations very close to the AP. On the other hand, Topology #4 is the topology least proportion of users located far away from the AP. Figure 6.15 depicts the total throughputs for all algorithms for the five topologies. The aggregate load is set as 200 Mbps, the maximum aggregate size is again 63. Our results indicate that the relative performances of all the algorithms are similar to previous results for the different topologies. Nevertheless, amount of performance enhancement depends on the exact topology. Specifically, AOS algorithm improves the network throughput by 15-26 % with respect to LQ, 35-52 % over MRS/PFQ and 53-94 % over SRPT in different topologies. CQS perform below AOS, as seen in the figure. The opportunistic algorithms perform better compared with LQ when the proportion of poorly located stations are higher since they can avoid transmitting over poorly conditioned stations. The MRS and SRPT algorithms are persistently lower than other opportunistic schedulers but particularly fails when few high capacity users exist, which leads to low aggregate sizes. Note that the OAR algorithms and the P-AOS algorithm are enhanced significantly in Topology #4. This is due to the increase of ratio of high capacity users, leading to higher aggregate sizes for the OAR algorithms and more transmission over high channel users for the P-AOS algorithm. In Topology #2, even though there are less users with good channels, there are users with similar conditions, leading to a higher aggregate size and higher throughput from most of the other topologies. P-AG and AOS may exceed each other depending on topology, with P-AG suffering in terms of throughput due to transmissions over users with poor channel conditions when the ratio of poor users is higher. P-WF consistently outperforms other algorithms since it can avoid low capacity users if necessary, but the least improvement is in Topology #2, since the aggregate sizes used by AOS are higher compared with other topologies since users are more evenly distributed. The throughput of the CQS algorithm is always closer to the throughput-maximizing schedulers but is always lower by about 5-10 % in terms of total throughput.

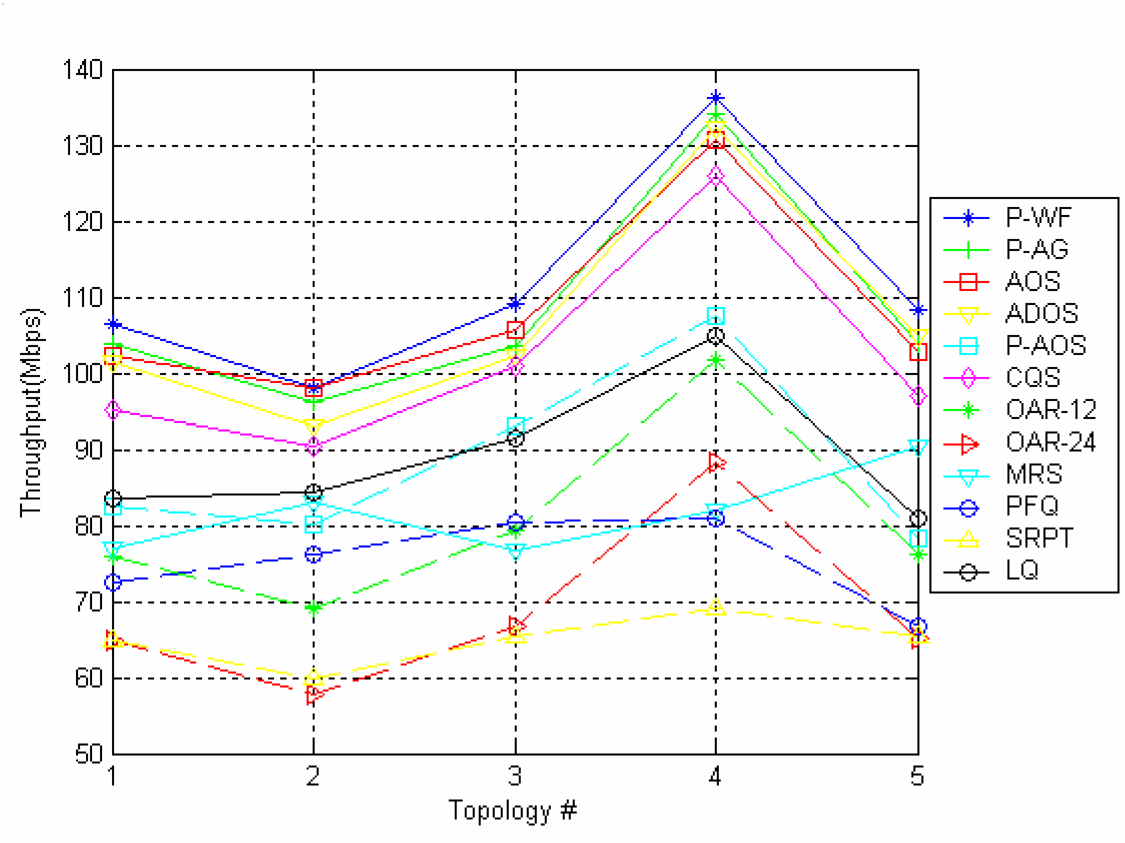


Figure 6.15 Performance of schedulers with different topologies.

The time averaged data rates are shown in Figure 6.16. Topology #4 yields the usage of the highest data rates since low capacity stations are less. On the contrary, the lowest average rates are used in Topology #2 due to the absence of very good located users. Note also that this also causes the average rates of the algorithms to be closest in Topology #2.

The average aggregate size over five topologies is shown in Figure 6.17. Parallel to the outcomes regarding average data rate, the highest aggregate sizes are present in Topology #2 since in order to gain access by maximizing the scheduling metric, the better positioned users have to increase their aggregate sizes further since their actual data rates are lower in Topology #2. The aggregate size of P-AG does not fluctuate to a great extent opposed to AOS, whose aggregate sizes vary with topology. Considering Figures 6.16 and 6.17, even though the data rates used in Topology #2 are significantly lower than other topologies, the throughput is not very low as compared to other topologies except Topology #4 since higher aggregate sizes yield better efficiency.

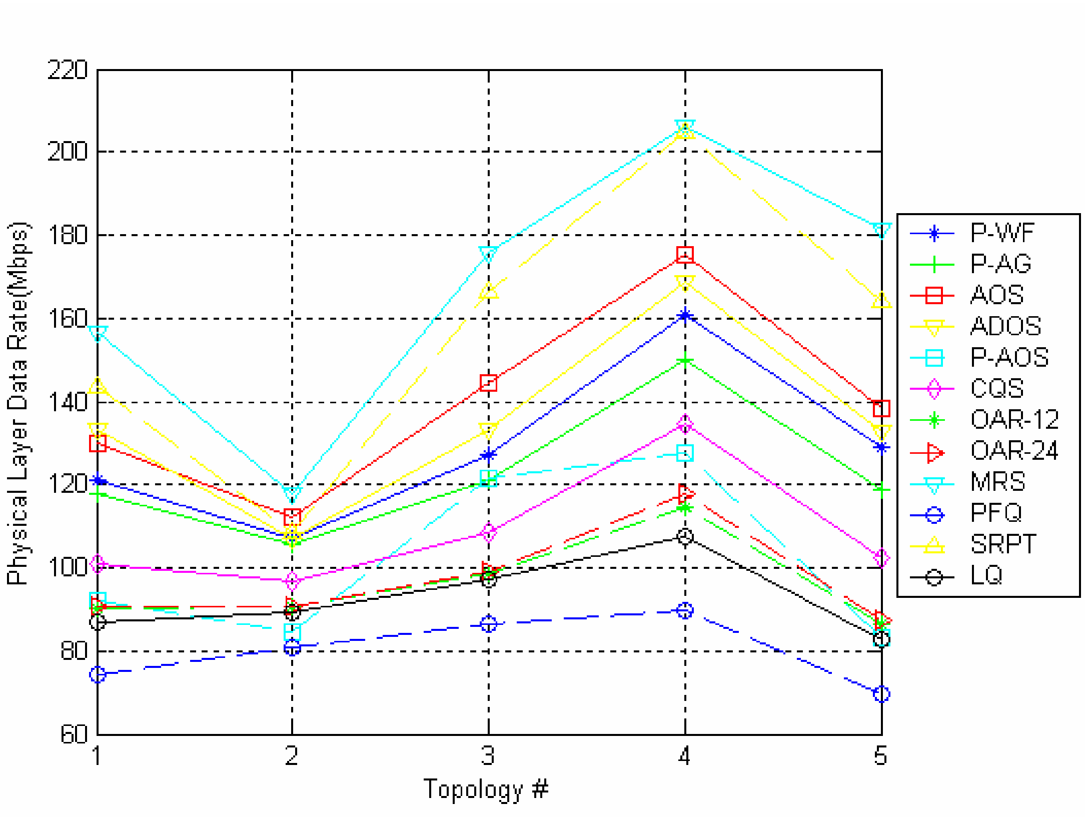


Figure 6.16 Average data rate depending on Toplology.

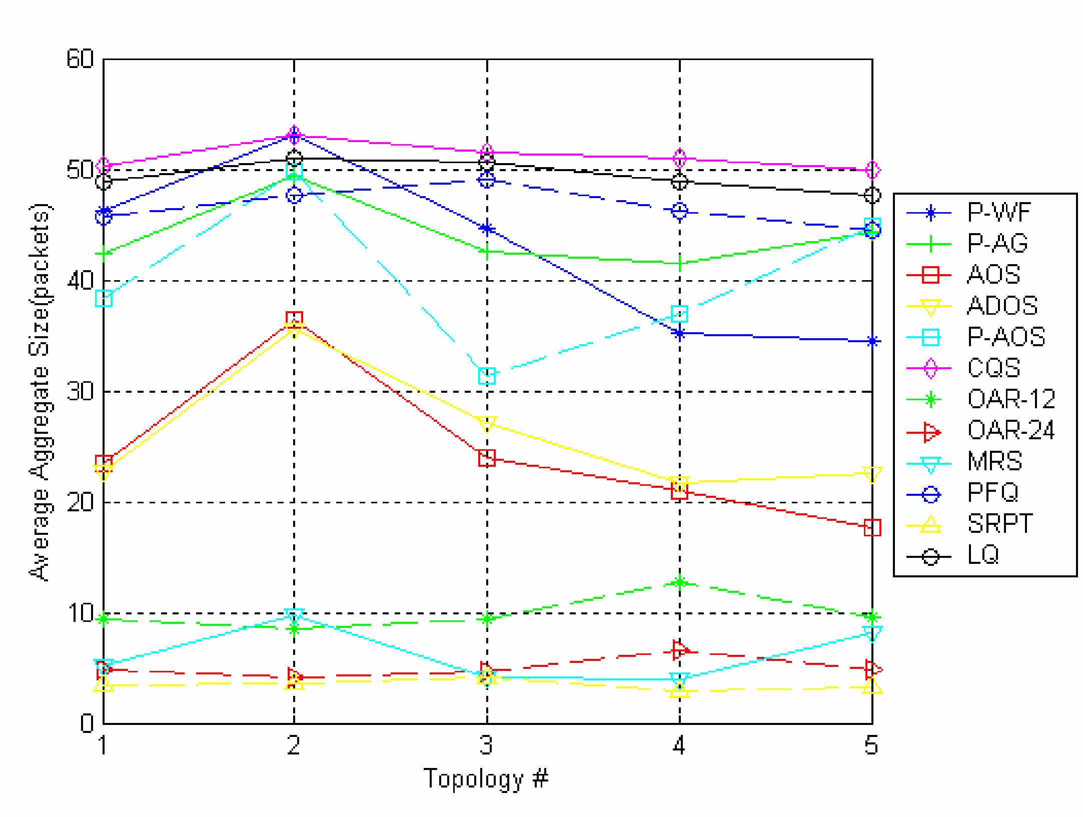


Figure 6.17 Average aggregate size depending on Topology.

Mean user delay is seen in Figure 6.18. Consistent with Figure 6.10, P-AG always yields the minimum average delay. The lower ratio of low capacity users in Topology #4 results in a lower average delay. The delays of AOS and P-WF are always greater than CQS and LQ but may exceed eachother depending on the Topology. CQS is slightly better than LQ in terms of delay. SRPT and MRS perform worst since some users are not served.

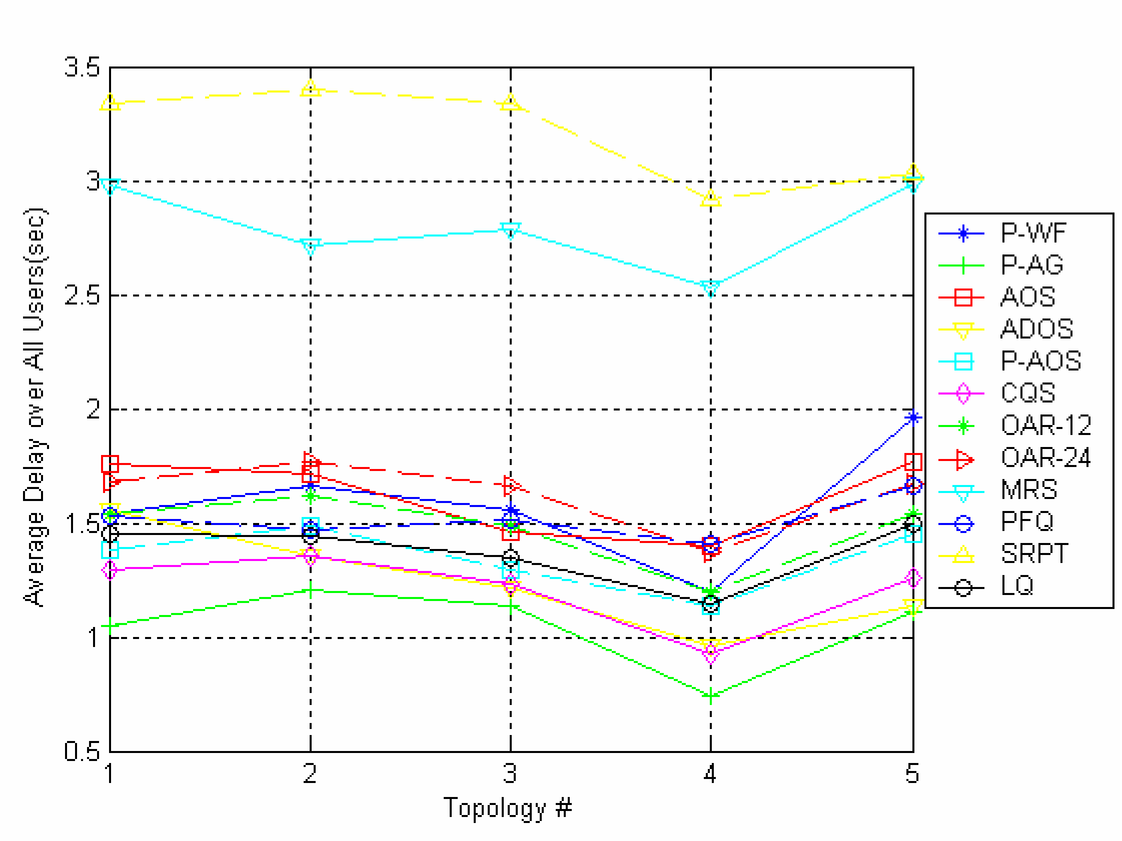


Figure 6.18 Mean user delay for different Topologies

Fairness performance depending on topology is seen is Figure 6.19. Even though the overall fairness performance is parallel for all topologies, Topology #4 performs best since the ratio of low capacity users is the lowest, yielding the lowest number of users with low throughput. P-AG is consistently fairer than AOS as expected due to access guarantees.

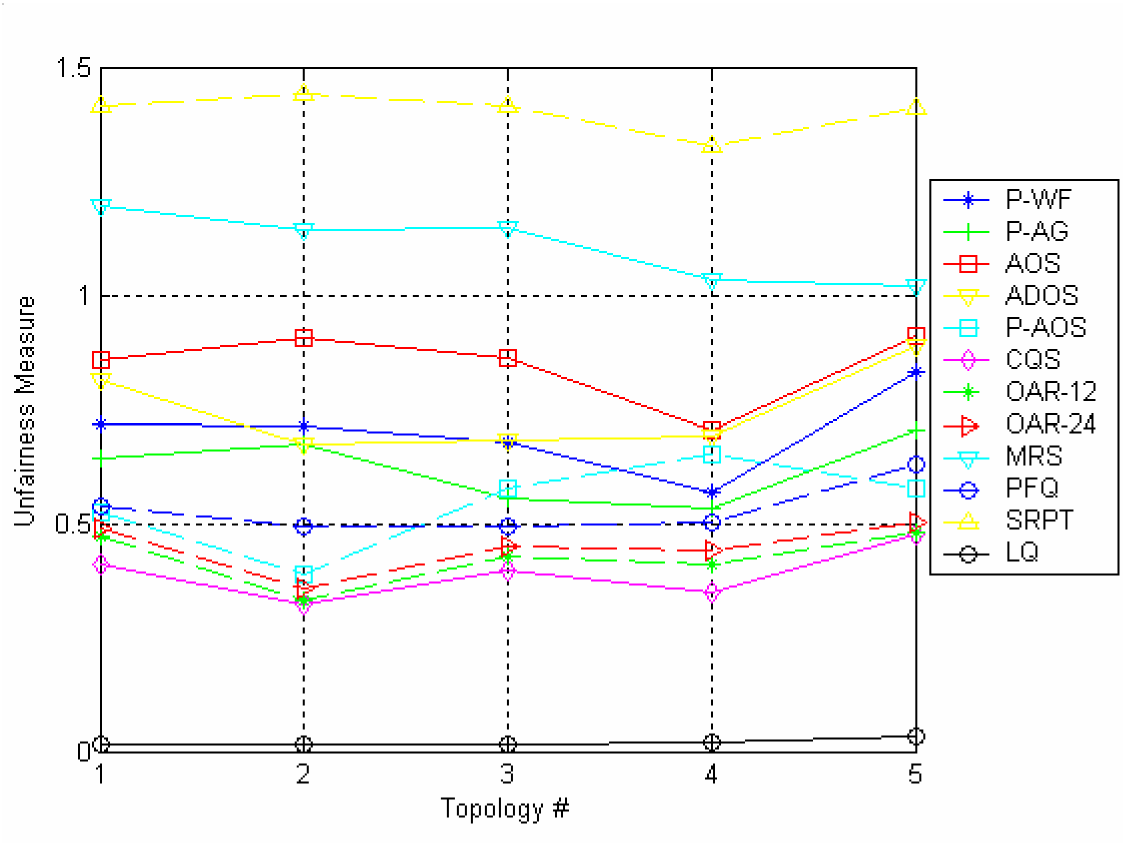


Figure 6.19 Fairness of schedulers with different topologies.

6.4.2 Performance with Varying Load

Next, we present our simulation results for varying load with different topologies. Simulations have been carried out in 10 different topologies consisting of 12 stations distributed within 25 m of the AP. The average values of throughput, data rate and aggregate sizes over these 10 topologies are demonstrated in Figures 6.20-6.22. Consistent with previous results, our proposed algorithms outperform existing algorithms. The algorithm which offers the highest throughput out of the existing algorithms is the LQ algorithm. Out of the queue aware algorithms, the AOS and ADOS algorithms exceed LQ by about 18 % and the MRS algorihm by 39 %. The CQS algorithm offers about 5 % lower throughput as compared with the AOS algorithm. Controlled access schedulers further enhance throughput. The P-WF algorithm offers the highest throughput out of all algorithms, exceeding the AOS algorithm by 4 % . The P-AG algorithm yields identical throughput values with the AOS algorithm while providing access guarantees to every user in the network.

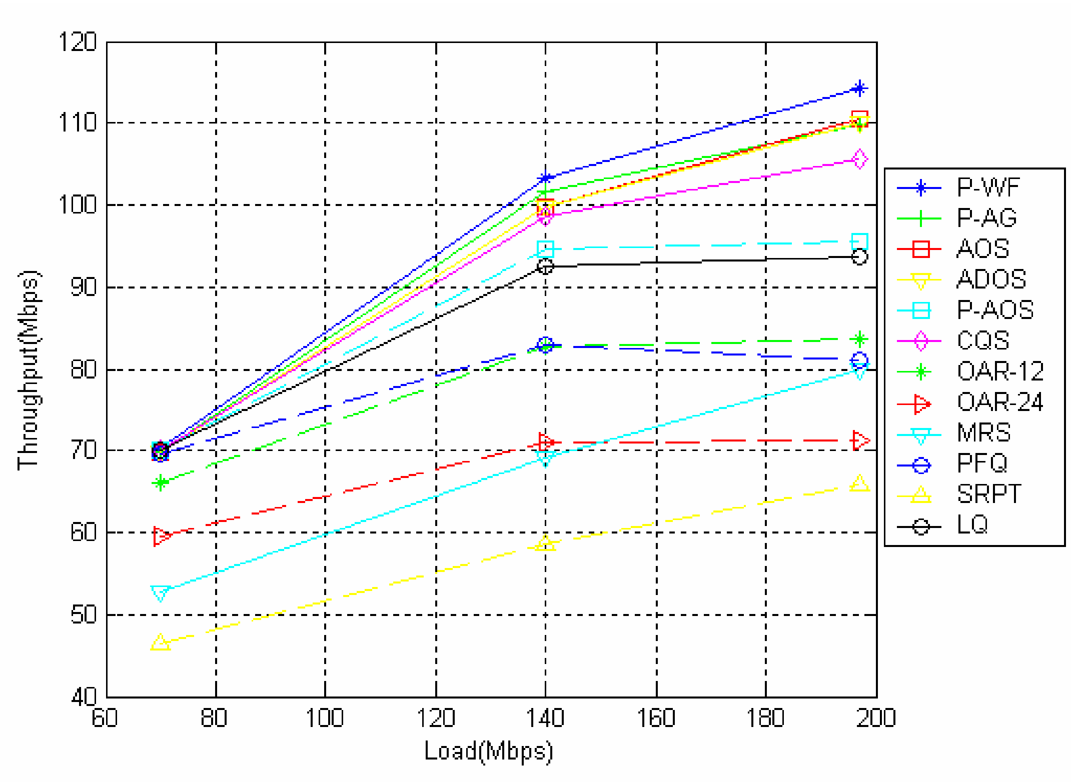


Figure 6.20 Throughput with varying load averaged over multiple topologies

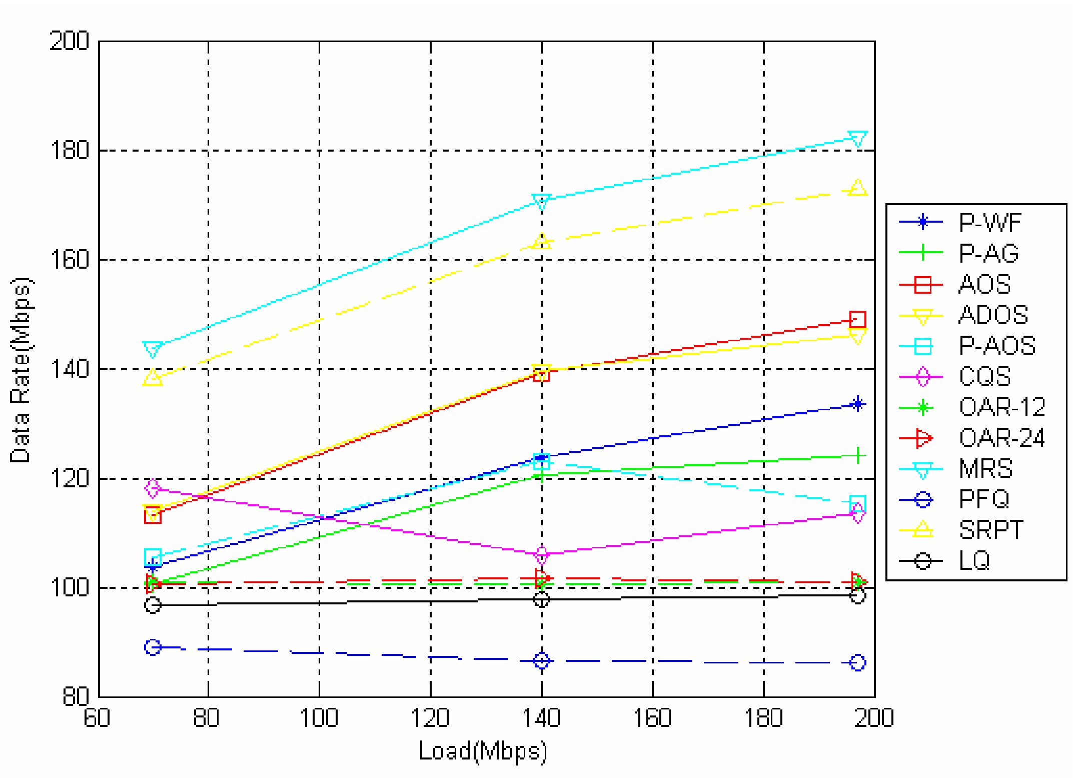


Figure 6.21 Data rate with varying load averaged over multiple topologies

In accordance with previous results, even though the MRS and SRPT algorithms operate over the highest data rates as seen in Figure 6.21, the resulting throughput values are low due to much lower aggregate sizes as compared to other algorithms as shown in Figure 6.22. The CQS and LQ algorithms provide the highest MAC efficiency as given in Figure 6.23, since these algorithms operate with the highest aggregate sizes.

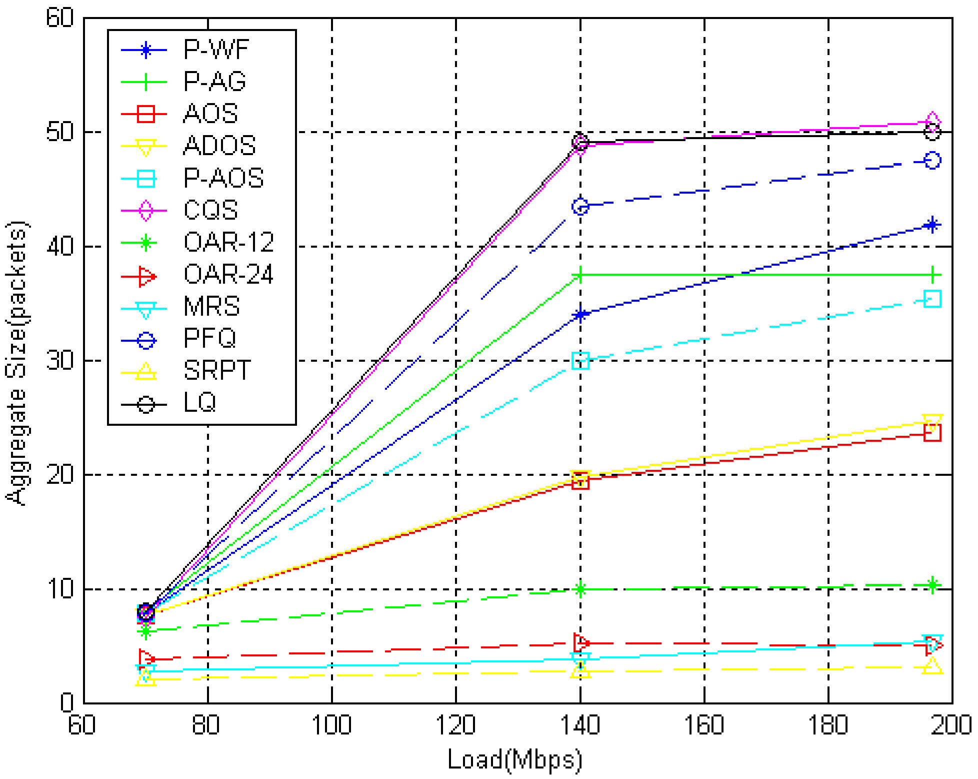


Figure 6.22 Aggregate Size with varying load averaged over multiple topologies

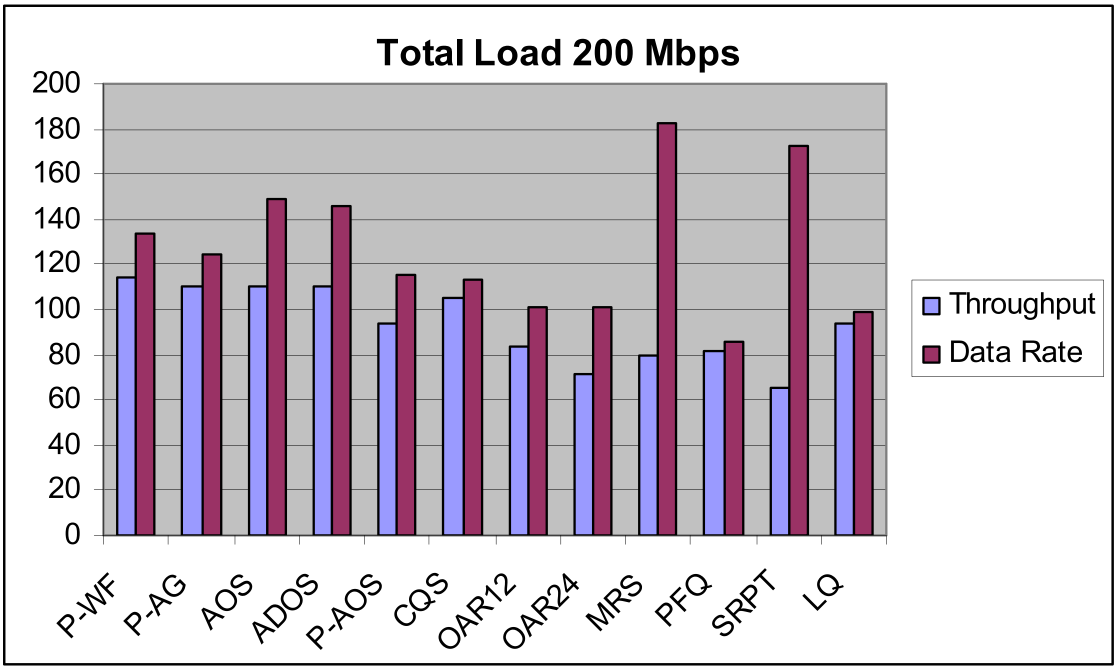


Figure 6.23 Throughput vs. Data Rate averaged over multiple topologies

6.5 Performance with Relaying

In this section we analyze the effect of incorporating relaying with opportunistic scheduling with frame aggregation. The AP calculates the effective *relaying rate* for each station by using the expression in (3.13). For a specific station, which we call *final station*, all other stations in the network are considered as possible intermediate relaying stations and the intermediate station which maximizes (3.13) is selected as the intermediate relaying station for the destination station. After determining the most suitable relaying station, the new relaying rate offered by using the intermediate station is compared with the *direct rate*, which is the rate used by directly transmitting from the AP to the destination station. If the direct rate is greater, then we decide that possible transimssion will be carried to that station without using relay stations.

After determining whether relaying is beneficial for a station and the relaying rate, the schedulers are implemented with the new rates used for the metric calculations of stations which require relaying. Typically, relaying will the rates of stations with poor channel conditions which are located far away from the AP, equivalently increasing their metrics, increasing their chances for being served by the AP. As a result, we expect relaying will improve fairness pefrormance of schedulers. In addition, since higher effective data rates are used, relaying shuold improve throughput of the non-opportunistic scheduler LQ. For opportunistic schedulers, both effective data rates are and the proportion of service for users with poor channels are expected to increase so we can not absolutely argue that relaying will enhance throughput.

Before presenting the results over the three simulation sets, we first present how the average supported data rates vary by distance. The average data rates are arithmetic averages, not time averages. Note that the values of the IEEE 802.11n data rates are also important. Apart from the direct average data rates, we also present the rate that would have been achieved if relaying was applied with two equal hops, neglecting the overhead factor in the real effective relaying rate. Here, we simply divided the rate for half of the direct distance by two. A point-to-point scenario of an AP and a station was used for the simulations.

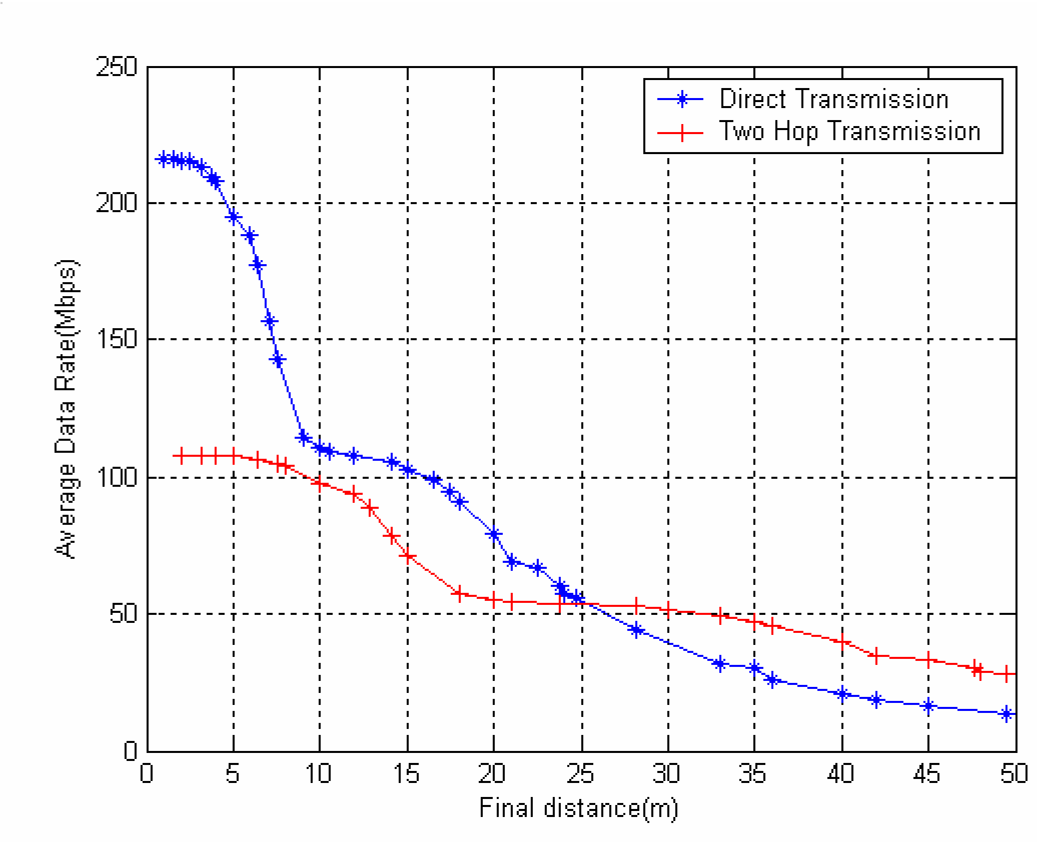


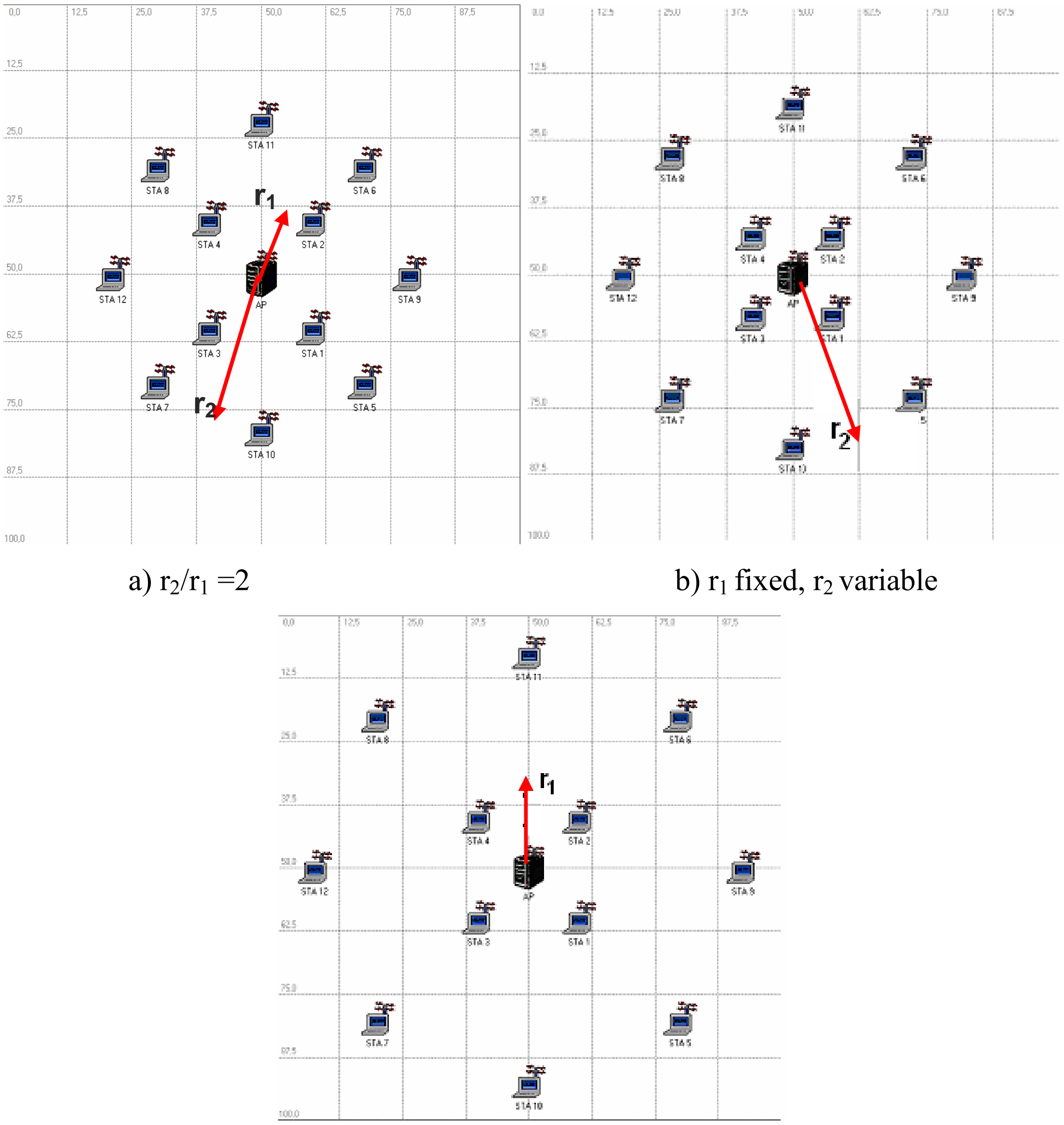
Figure 6.24 Comparison of data rates for direct and two-hop neglecing overhead

Since the maximum data rate is limited, relaying does not offer improvement for short distances since the maximum data rates are already realized by transmitting over one hop. On the other hand, as distance is increased, the direct transmission rate reduces significantly. Transmitting over two hops offers significant increase in the effective average data rates when the direct distance is high since the data rates of the individual hops have not encountered the region where the data rate is decreased significantly. As a result, we should expect that the throughput performance will be improved if relaying is used for users far from the AP. However, we should expect that relaying would not be advantageous for very high distances since the data rates of individual hops would also be at the lowest rates.

Obviously topology has a strong influence on the relaying decisions and the possible gain offered by relaying. Distances between the AP and destination stations, AP and intermediate stations, and intermediate stations to destination stations define the respective data rates to be supported in between, so we analyzed relaying in three types of network structures by varying the distances of all or a subset of the stations. The networks analyzed from 6.1- 6.4 do not consist of stations which will be significantly enhanced via relaying, so we consider larger networks for the analysis of relaying. Our networks used for relay performance evaluation consist of 12 stations and the AP. We divide these stations into three subsets: Four inner stations that are closest to the AP seperated by a distance r_1, which is the *inner radius* of the network. We denote this set of stations by ***A***. The remaining eight stations are r_2 meters away from the AP, which is the *outer radius* of the netwok. These outer stations can be further defined to two subsets: Four of them denoted by set ***B*** are alligned with the inner stations, which minimizes the distance from the related inner station, increasing the likelihood that a relatively higher data rate can be supported from the inner station to the outer station. The rest of the stations are defined by ***C*** are not alligned with the inner stations, yielding a higher distance between the inner stations. The alligned stations are closer to the inner stations than the non-alligned nodes, so it is expected that the data rates between the respective elements of ***A*** and ***B*** will be higher than the rates between the elements of ***A*** and ***C.*** Consequently, the probability of deciding in favor of relaying for stations in ***B*** is higher as compared to stations of ***C.*** Moreover, with a similar reasoning, the relaying rates for ***B*** are likely to be greater than the relaying rates of ***C***. Since all the schedulars except LQ use capacity or rate information in the scheduling instant, this differentiation due to relaying varies the stations selected. We expect that after relaying, the stations in ***B*** may gain higher access proportions as compared to ***C*.** Nevertheless, the stations in the outer radius are expected to be served more frequently.

The actual distances are very important for the impact of relaying. Accordingly, in order to analyze the performance of schedulers with relaying we varied the inner radius, r_1 and outer radius, r_2. The total load was 200 Mbps evenly distributed between the users. The algotithms AOS, CQS and LQ are compared with their counterparts aided by relaying. Basic rate was selected as 12 Mbps for the simulations.

In the first simulation set, both inner radius r_1 and outer radius r_2 are varied, with the ratio of the distances set to 2. One realization can be shown in Figure 6.25a . Accordingly, both the good and bad positioned stations are effected in a similar manner in terms of the increasing or decreasing of supported data rates. Yet, the fact that they both increase or decrease does not imply that their ratio remains fixed. In the second set shown in Figure 6.25b the inner radius r_1 is fixed and the outer radius r_2 is varied. Thus, the supported data rates between the AP and ***B***, ***C*** and ***A*** and ***B*,*C*** are varied, affecting both the relaying and the scheduling decision.



a) r_2/r_1 =2

b) r_1 fixed, r_2 variable

c) r_1 variable, r_2 fixed

Figure 6.25 Relaying Topologies

Finally, in the set shown in Fig 6.25c as an instance, the outer radius is fixed and the inner radius is varied. This yields that the supported data rates between the AP and ***A*** and ***A*** and ***B***, ***C*** are varied, again changing both the realying decision and scheduling decision as in the second simulation set.

Figure 6.26 demonstrates the results belonging to simulation set #1. As expected, total throughput decreases with increasing "network radius" since the supported data rates are likely to decrease for all stations. However, the effect of distance on total throughput largely depends on the scheduling algorithm in concern. When the radius is small, the algorithms perform very similar since the topology is close to a uniform topology. As radius increases, AOS and CQS outperform LQ better positioned stations are frequently preffered more. LQ yields very low throughputs at over large distances to further stations since the algorithm can not avoid transmitting to the outer stations, even if the supported data rates are very low. Another reason of the extremely low throughput of LQ is that even if the arithmetic averages of the data rates used is much higher than the total throughput, the effect of bad positioned stations dominates the overall performance. This result is due to the fact although that LQ serves each user equally in terms of data, the actual temporal shares of each user are significantly different. Users with very low transmission rates are served a very long duration, reducing total network throughput drastically.

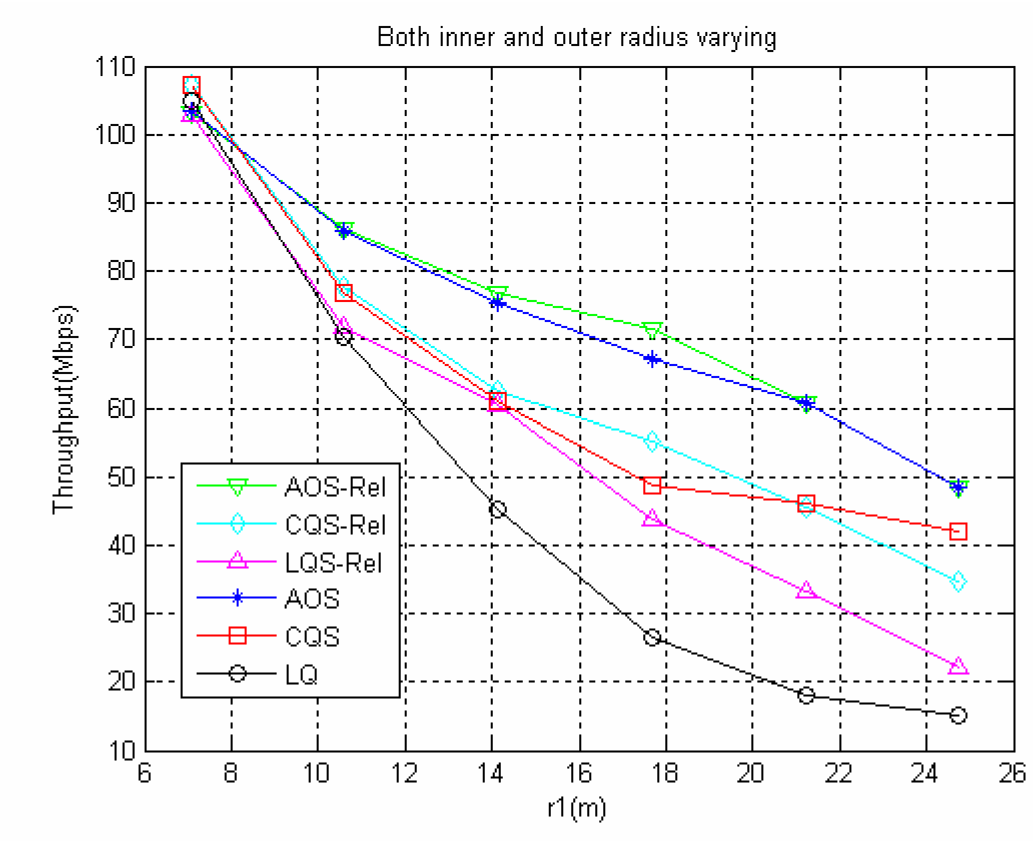


Figure 6.26 Throughput with both inner and outer radius varying

Among all schedulers, the LQ algorithm ,which is the algorithm which transmits mostly to bad conditioned stations benefits most from relaying. The effective rates which the bad conditioned stations are served are increased, leading to significant increase in throughput.

The performance and overall behaviour of the AOS algoritms is not varied notably due to relaying. For small networks, relaying is not used since either because it is not favored due to short distances, or the stations which can be enhanced through relaying are selected for transmission when their individual instantaneous rates are high. When the distance is increased, initially we see that relaying is employed since the distances are increased, and the probability of deciding for relaying is increased. However, as the distance is further increased, relaying is not exploited at all. This is due to the fact that as the network is enlarged, the inner stations also transmit using lower rates, preventing goodly positioned stations to fully serve their stations. Since the inner stations are not fully served, their scheduling metrics stay much higher than the lower stations, resulting in allocating the resources always to inner stations, without transmitting to outer stations.

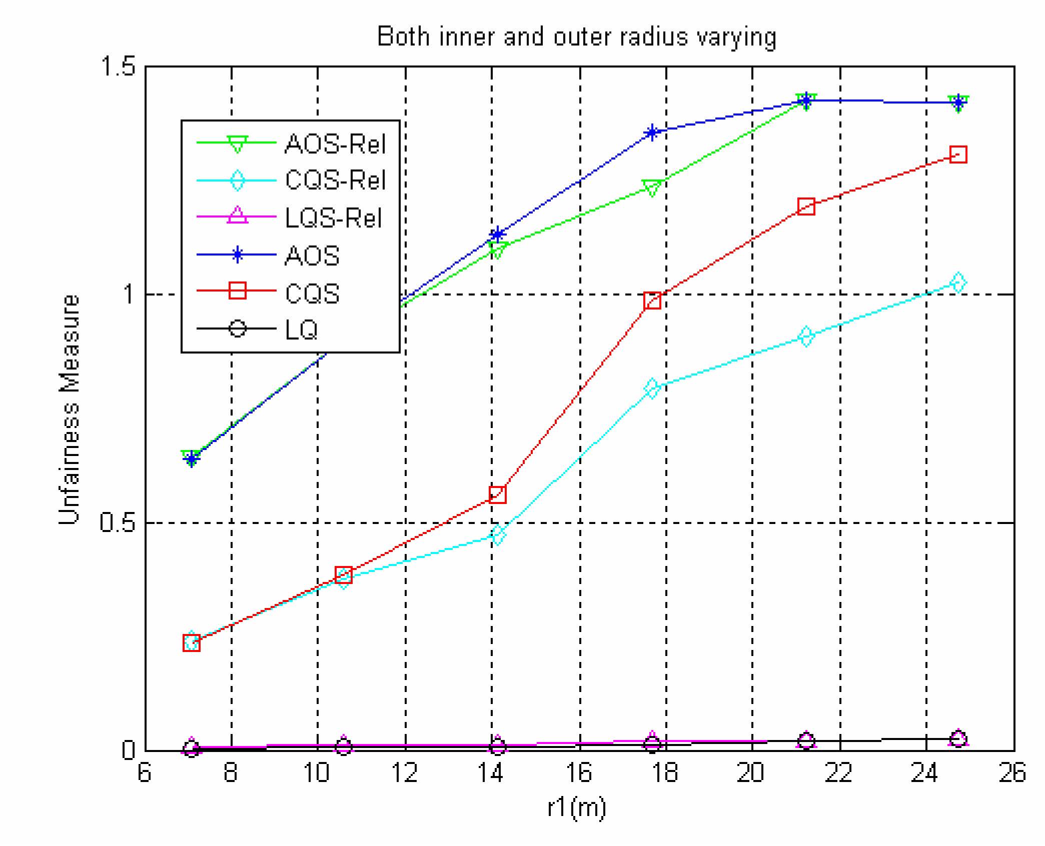


Figure 6.27 Unfairness with both inner and outer radius varying

The fairness is slightly improved when relaying is employed for AOS since the chance of gaining access inreases especially for the users in ***B*** as shown in Fig 6.24.

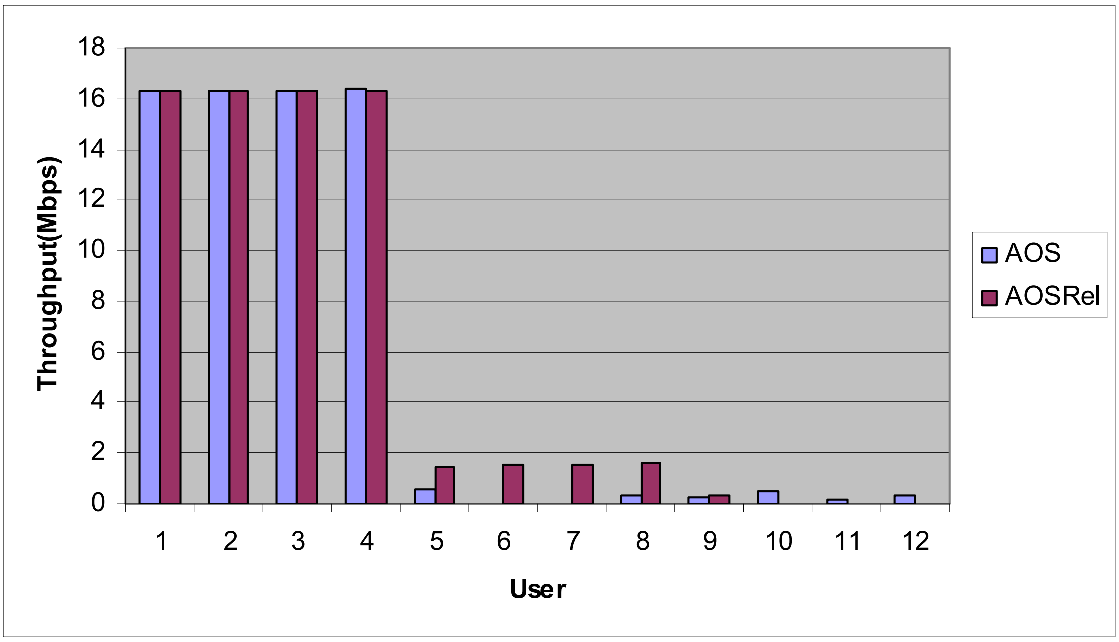


Figure 6.28 Individual user throughputs for varying radius (r_1=17.67)

The behaviour of the CQS algorithm is different than both LQ and AOS. Relaying is employed after a distance. The frequency of employing relaying is not as high as LQ since better stations are favored and outer stations may be selected when their own channel coditions are good without the need of relaying, but the proportion of transmitting with relaying is still much higher compared with AOS. We see that relaying significantly improves fairness preformance since it yields increase in the capacity term in the relaying metric, enabling the outer stations to gain access without growing their queue sizes as much as the case without relaying. An exapmle of individual user throughputs can be seen in Fig 6.29. However, the impact of relaying on throughput for CQS is not always positive. Initially relaying offers an improvement for throughput but afterwards for high distances the CQS algorithm realizes a decrease in throughput due to the improvemet on fairness. The inner stations are selected with a lower proportion compared with the case without relaying, and for very large distances where the outer stations are away this causes the throughput to slighlty decrease.

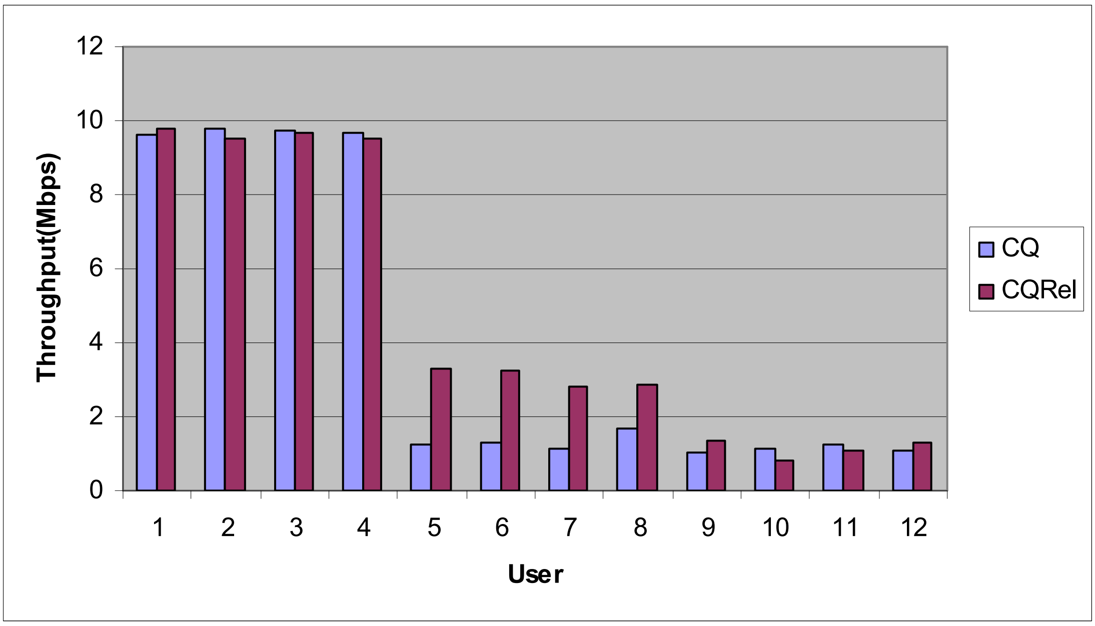


Figure 6.29. Individual user throughputs for varying radius (r_1=17.67)

Next, the results belonging to the simulation set #2 where the inner radius is fixed and outer radius is varied is presented in Figures 6.30 and 6.31. As in the simulation set #1, the algorithms perform similar when the distances are small and the network is close to uniform. Relaying enhances throughput for LQ the most as in simulation set #1 since the outer stations are selected frequently, with high temporal access share. Note that the advantage due to relaying is maximum at intermediate distances since the intermediate distances between the inner and outer stations are in a region where the data rate decreases slower than the direct hops. However, when distance is further increased, the data rates supported by the intermediate hops also reduce significantly and the advantage is slightly reduced. Yet, relaying offers improvement since when the two intermediate rates start to differ significantly, the overall relaying rate in (3.13) is dominated by the lower rate and the difference between the direct rate and intermediate rates from the inner to outer stations still enables relaying to increase throughput.

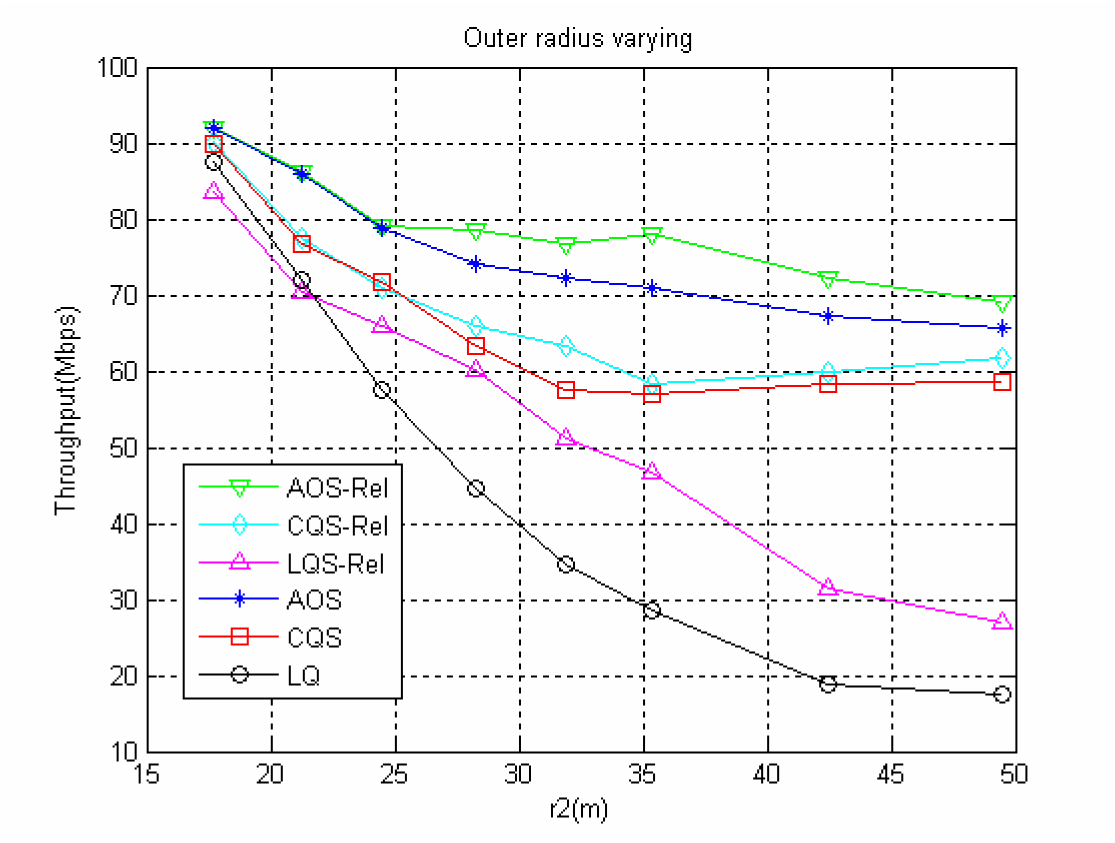


Figure 6.30 Throughput with outer radius varying

Opposed to simulation set #1, relaying is used in the AOS algorithm except small networks for all distances and the throughput is improved to a larger extent. The reason of why relaying is used is that since the inner stations are placed close to the AP and relatively high data rates are used for the inner stations. For all simulations the AP is able to serve the inner stations without filling their queues excessively, enabling also the outer stations to gain access since the metrics of outer stations can exceed the metrics of inner stations when the queue sizes of the inner size stations are small. The overall total throughput is slightly decreased for larger r_2 since the outer stations are more far away and the relaying rates also decrease. The best improvement is for intermediate outer distance radius values, as in the case of LQ.

The CQS algorithm offers slight improvement in throughput in addition to significant improvement in fairness. Like LQ and AOS, the best improvement is offered for intermediate outer radius distances. After some r_2 value , we see that the throughput increases with and without relaying as opposed to LQ and AOS. This increase can be attiributed to the fact that as the outer radius increases too much the rates supported by transmitting to outer stations decrease , causing them to get much less access proportion at lower outer radius distances since the inner data ratres remain high. Increasing the access proportion for inner users prevails over the decrease in rate while transmitting directly to outer stations, yielding increase in thoughput. The difference between the CQS and AOS algorithms in this perspective is that AOS already eploits the inner stations as much as possible but CQS does not serve all of the data to the inner stations, yielding room to improvement.

As seen in Figure 6.31 , the unfairness measure increases for the AOS and CQS algorithms regardless of whether relaying is employed as the outer radius is increased, since the ratio of the capacities of the inner to outer stations increases, effecting the scheduling metrics and enabling inner stations to gain access with lower queue sizes compared with lower r_2.

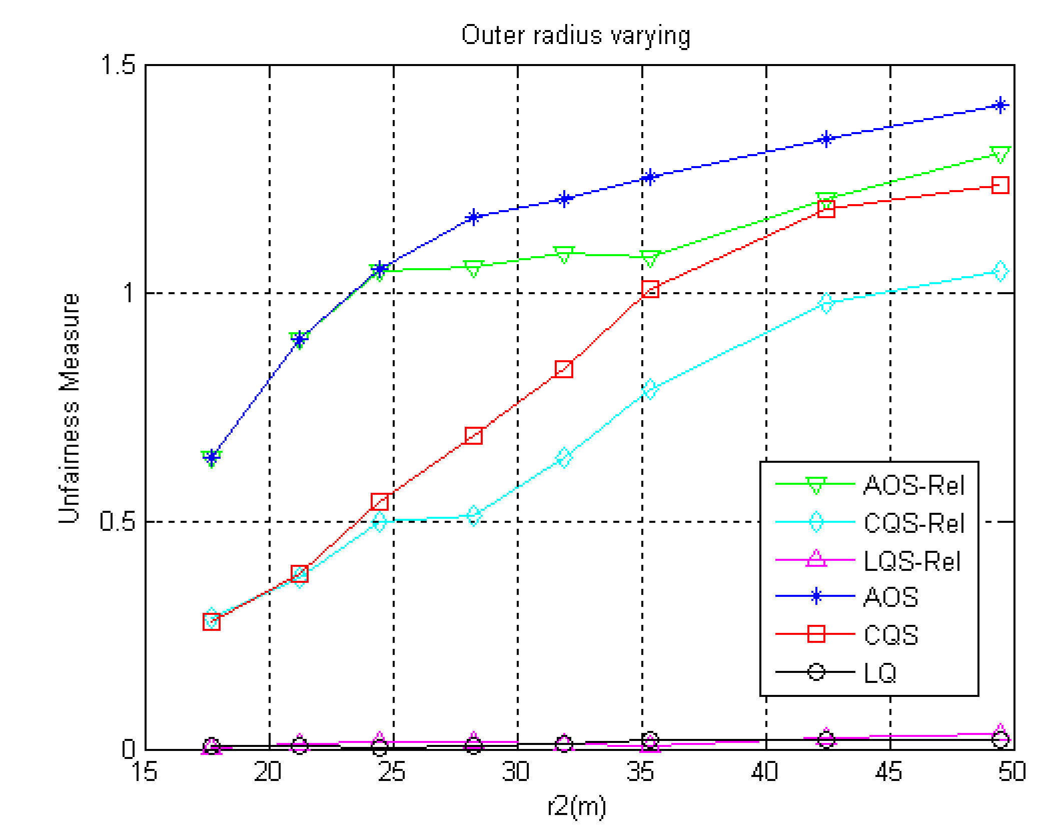


Figure 6.31 Unfairness measure with outer radius varying

Finally, the results of the simulation set #3 where the outer radius is fixed and the inner radius r_I is varied is presented in Fig 6.32 and 6.33. When inner stations are close to the AP, we see that the throughput of LQ with relaying is low compared with intermediate distances. This result may seem unexpected at first glance since the data rates of inner stations reduce but when the inner stations are close to the AP, the distances between the inner and outer stations are high, leading to low transmission rates even in the presence of relaying. Low rate stations dominate the overall throughput due to long transmission durations, and improving the effective relaying rates prevails over using hişgher data rates for the inner users. After further increasing the inner radius, the relaying rates also reduce significantly since the distances form the AP to inner stations decrease, leading to very low throughput values even though the inner and outer stations are close.

AOS initially uses relaying when the inner radius is low since the inner stations are served with a high data rate and the outer stations can also gain access. Relaying offers higher rates for outer stations which are placed far away from the AP and improvement in both fairness and throughput is achieved. However, as r_I is increased, the inner stations cannot be served very frequently due to lower data rates, causing the metrics of inner stations to always exceed the scheduling metrics of outer stations. Hence, relaying is not applied, also due to the fact that offered relaying rates are not advantageous compared with the relaying rates of lower inner radius values. The throughput values are very low due to the fact that inner stations also use low data rates for very high inner radius values.

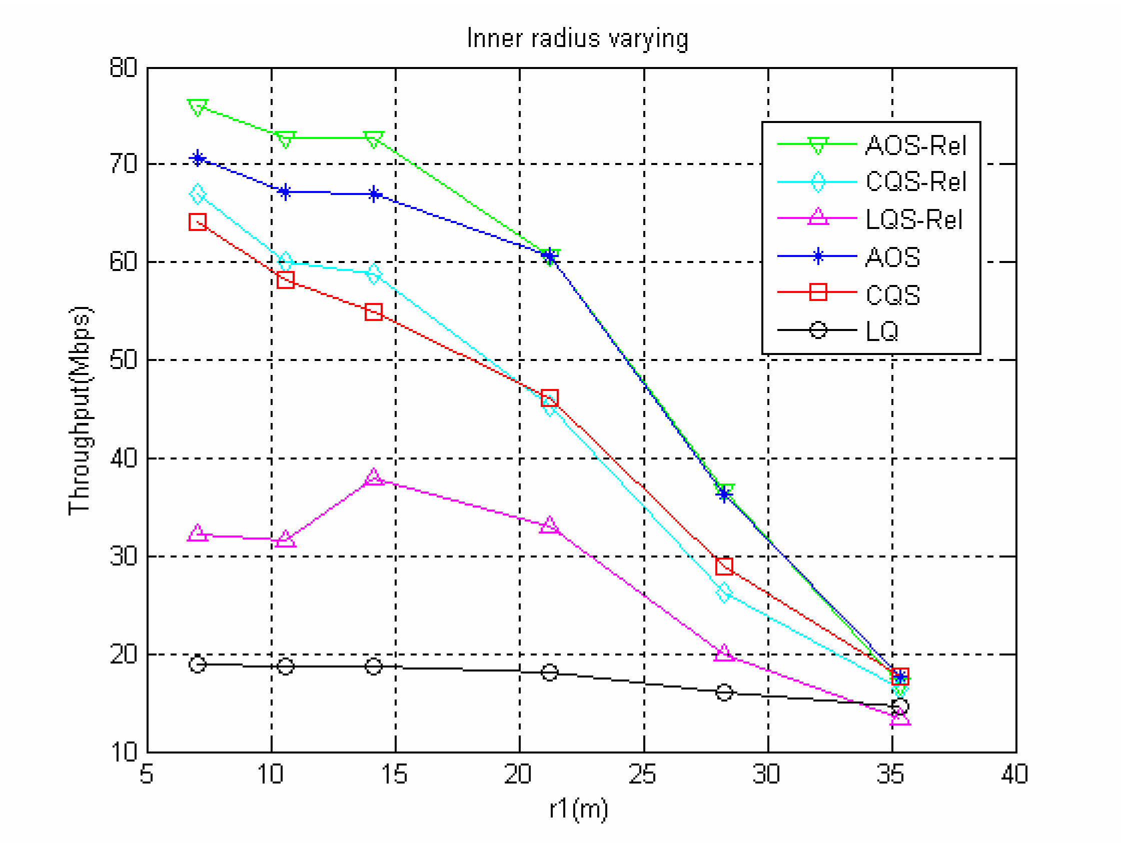


Figure 6.32 Throughput with inner radius varying

Throughput performance of CQS is not effected severely by the presence of relaying. For intermediate inner radius values which are similar to the previous simulation sets, throughput is slightly enhanced. However, when inner distance is increased the throughput is slightly decreased since the data rates are low even with relaying and improving fairness results in a reduction in throughput. The unfairness measure is significantly lower with relaying for CQS.

For all algorithms, we see that the unfairness measure decreases with increasing inner radius value for high inner radius, except for AOS, which unfairness measure increases for intermediate values. This is due to the fact that there is still significant difference between the inner and outer stations and the inner stations can not be served frequently as they were for low r_1 values, leading to less access for outer stations. In contrast, for low distances the inner stations are served, yielding to access for other stations and for high distances the capacities of inner and outer stations become closer.

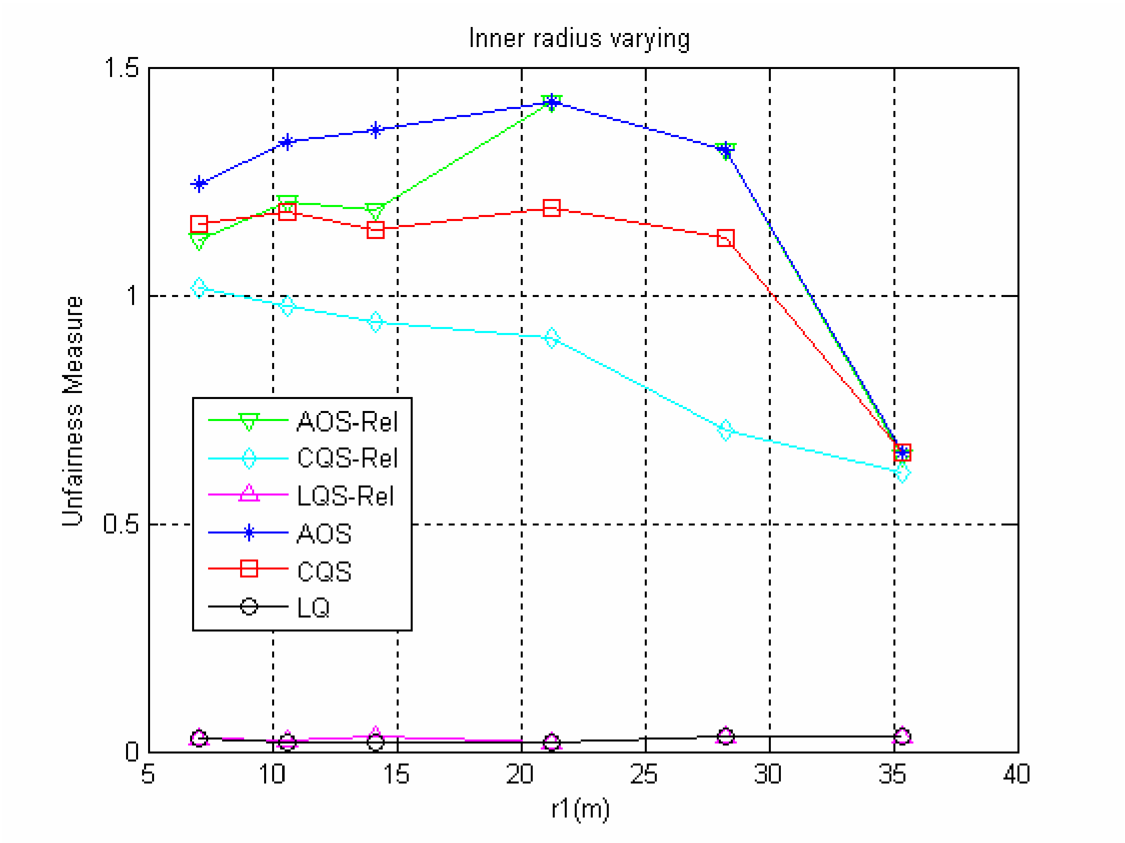


Figure 6.33 Unfairness measure with inner radius varying

CHAPTER 7

CONCLUSIONS

7.1 Contributions

In this thesis, we have proposed new scheduling algorithms that consider both channel and queue states for use selection in the next generation WLANs, namely the IEEE 802.11n standard, which employ MIMO technology and frame aggregation. Through detailed simulations, we have shown that with frame aggregation, spatially greedy scheduling algorithms are not optimal in terms of throughput. Even though the data rates used are the maximum over all algorithms and they would have provided the highest throughput values in the presence of overhead, they fail considerably under the 802.11n model.

Our new queue aware scheduling algorithms, AOS, ADOS and CQS, show significant increase in total system throughput, up to 53 %, as compared to opportunistic schedulers, MRS and PFQ. With aggregation, the simplest non-opportunistic method, namely LQ, performs better than MRS and PFQ, but falls behind our schedulers. The OAR algorithm, which was proposed considering aggregation transmission, falls below other schedulers in terms of throughput, mainly due to low aggregate sizes. Our algorithms offer a compromise between throughput and fairness. AOS and ADOS offer better throughput enhancement, while CQS is more suitable for situations where fairness is more important. P-AOS, which is the counterpart of the PFQ algorithm improves throughput compared with the PFQ algorithm. Last but not least, performance enhancement depends on the network topology and geographical distribution of the stations.

Our access controlled schedulers which maximize throughput over longer time scales further enhance system throughput as compared with our queue aware schedulers, justifying the concept that selecting the user which maximizes the instantaneous scheduling metric may not provide maximum performance throughout the entire time duration. The P-AG algorithm offers better fairness and moreover QoS guarantees as opposed to the AOS algorithm while providing similar throughput. Even though the P-WF algorithm does not provide access guarantees due to the nature of the Waterfilling algorithm, the offered throughput is consitently maximum out of all algorithms, exceeding AOS by about 4 %.

When average user delay is examined, more opportunistic schedulers such as POS and SRPT fail since they do not serve a cosiderably large amount of users. AOS and P-WF provide lower delay but the average delay is slightly greater than fair schedulers such as CQS and LQ. In addition to providing access guarantees at high thoughput values, the P-AG algorithm always provides the lowest mean user delay.

Applying the concept of relaying is slightly differentiated from convential relaying due to overhead, and we have shown that for networks which have users located far away from the AP ,relaying improves either throughput or fairness, or even both simultaneously. Our queue aware schedulers are not improved through relaying in terms of throughput as much as non-opportunsitic schedulers since poor channel users are not selected frequently, yet performance is enhanced.

7.2 Future Work

Channel states were assumed to be available instantaneously at the AP in the analysis throughout this work. The delay due to accurate channel information may slightly reduce system throughput. Nevertheless, the effect of this additional delay is present for all schedulers except the LQ algorithm, which is the only algorithm which does not take channel information into account. Hence, proposed algorithms will not suffer more as compared to existing opportunistic scheduling algorithms when channel delay is present.

Even though network throughput is largely determined by downlink scheduling for infrastructure based WLANs, scheduling for system throughput maximization for uplink transmission systems may also be developed using similar analysis in this thesis. Since data is first transmitted to the AP and then forwarded to the destination station, a user to transmit data for more than one users may determine the user to transmit using the relaying extension to our algorithms where the first hop is the one between the source station and the AP, and the second hop is between the AP and the destination station. However, implementation would be more complex compared to downlink since the source station requires channel information between the AP and destinations, which may prove too complex for a user as opposed to an AP. Uplink transmission to maximize throughput may also be carried out based on principles of downlink schedulers by adjusting the allowed reverse transmission duration from the AP for the source users. The AP may grant reverse duration in a fashion parallel to algorithms such as AOS, ADOS, P-AG and P-WF provided it has some insight on the individual user traffic or queue states.

In addition to single-destination aggregation systems, the principle of scheduling by considering channel and queue states can be applied to multiple- destination aggregation systems. Forming the group by selecting the common rate as the highest data rate supported by the stations may not be optimal in terms of throughput, since selecting a lower common rate may lead to a higher throughput since the overall block may result in a higher aggregate size

Finally, another extension planned is to exploit the possibility of serving multiple users simultaneously. Note that only one user is served in the current framework with multiple independent data streams transmitted for the destination station in the case of MIMO transmission. By forming new equivalent channel matrices by merging rows associated with different receivers, higher channel capacities may be offered. Multiple data streams for different users may be transmitted and decoded by the knowledge of channel matrices. However, the scheduling is also more complex for such a case since the transmission duration of the data streams must be very similar. Nevertheless, in addition to the downlink schedulers proposed in this thesis, the AP may examine if higher throughput is offered by transmitting simultaneously to different users.

REFERENCES

[1] S. A. Mujtaba, “TGn Sync Proposal Technical Specification 3,” *TGn Sync Technical Proposal R00*,August 13, 2004.

[2] I. Tinnirello and S. Choi, "Efficiency Analysis of Burst Transmissions with Block ACK in Contention-Based 802.11e WLANs," *in Proc. IEEE ICC'2005*, Seoul, Korea, May 16-20, 2005.

[3] C. Liu and A. P. Stephens.; “An analytic model for infrastructure WLAN capacity with bidirectional frame aggregation,” *Wireless Communications and Networking Conference, 2005 IEEE Volume 1*, 13-17 March 2005 Page(s):113 - 119 Vol. 1.

[4] R. Knopp and P. Humblet ,“Information capacity and Power Control in Single Cell Multi-user Communications,” *Proc. IEEE ICC’95*, Seattle, USA, vol. 1., pp. 331-335, June 1995.

[5] P. Viswanath, D. Tse, and R. Laroia, “Opportunistic beamforming using dumb antennas,” *IEEE Transactions On Information Theory,* Vol.48, No. 6, Page(s): 1277-1294, June,2002.

[6] X. Liu, E. K. Chong, and N.B. Shroff, “Opportunistic transmission scheduling with resource sharing constraints in wireless networks,” *IEEE Journal on Selected Area in Communications,* vol. 19, no. 10, pp. 2053-2064, Oct. 2001.

[7] A. Jalali, R. Padovani, R. Pankaj, “Data Throughput of CDMA-HDR: A High Efficiency-High Data Rate Personal Communication Wireless System,” *Proc. IEEE VTC Spring ’00*, Tokyo, Japan, pp. 1854-1858, May 2000.

[8] L.E. Schrage and L. W. Miller, “ The queue M/G/1 with the shortest remaining processing time discipline,” *Operations Research*, vol. 14, pp. 670-684, 1966.

[9] B. Sadeghi, V. Kanodia, A. Sabharwal, and E.Knightly, "Opportunistic Media Access for Multirate Ad Hoc Networks," *in Proceedings of ACM MOBICOM 2002*, Atlanta, GA, September,2002

[10] Kleinrock, L; “Queuing Systems , Volume I:Theory”, Wiley-Interscience, 1975

[11] Pahlavan,K; Krishnamurthy,P; “Priniples of Wireless Networks: A Unified Approach”, Prentice Hall, 2002

[12] IEEE Standards Department, “Wireless LAN Medium access control (MAC) and Physical layer (PHY) specifications”, *IEEE Standard 802.11-1997*

[13] Leon Garcia, A; Widjaja, I; “Communication Networks: Fundamental Concepts and Key Architectures”, Second Edition, McGraw Hill, 2004

[14] Y. Kim, S. Choi, K. Jang, H. Hwang, “Throughput Enhancement of IEEE 802.11 WLAN via Frame Aggregation”, *Proc. IEEE VTC Spring ’04*

[15] Paulraj,A; Nabar,R; Gore,D; “Introduction to Space-Time Wireless Communications”, Cambridge University Press, 2003

[16] Heath Jr., R.W. ; Love,D.J.; “Multi-Mode Antenna Selection for Spatial Multiplexing Systems with Linear Receivers”, *Proc. of the“ Allerton Conferenc on Communications Control and Computers”,* pp. 685-694, Monicello, IL, Oct. 1- 3, 2003

[17] E.Telatar, “Capacity of multi-antenna Gaussian channels,” *AT&T Bell labs Tech. Memo.,* June 1995.

[18] Haykin,S; Moher,M; “Modern Wireless Communications”, Pearson Prentice Hall, 2005

[19] H. Bolcskei, D. Gesbert and A. J. Paulraj ,“On the Capacity of OFDM-Based Spatial Multiplexing Systems,” *IEEE Transactions on Communications*, no. 2, Feb 2002 pp. 225-234.

[20] J. Wang, H. Zhai, Y. Fang and M.C. Yuang, ``Opportunistic media access control and rate adaptation for wireless ad hoc networks," *IEEE International Conference on Communications (ICC'04)*, Paris, France, June 2004.

[21] E.N. Çiftçioğlu, Ö. Gürbüz ,“Opportunistic Scheduling with Frame Aggregation for Next Genaration Wireless LANs,” *IEEE International Conference on Communications (ICC) 2006*, *Istanbul, Turkey,* June 2006

[22] J.Boyer, D.D. Falconer and H.Yanikomeroglu, “Multihop diversity in wireless relaying Channels”, *IEEE Transactions on Communications,* vol. 52, no. 10, pp. 1820-1830, October 2004

[23] V. Sreng, H.Yanikomeroglu, and D.D. Falconer, “Capacity enhancement through two-hop relaying in cellular radio systems”, *IEEE Wireless Communications and Networking Conference(WCNC ‘02),* 17-21 March, 2002, Orlando, FL, USA.

[24] J. Bicket, D. Aguayo, S. Biswas, R. Morris, “Architecture and Evaluation of an Unplanned 802.11b Mesh Network”, *MobiCom’05*, August 28–September 2, 2005, Cologne, Germany.

[25] V. Navda, A. Kashyap ,S. R. Das “Design and Evaluation of iMesh: an Infrastructure-mode Wireless Mesh Network”, *IEEE Symposium on a World of Wireless, Mobile ans Multimedia(WOWMOM),* Italy, June 2005

[26] OPNET Technologies, Inc^{TM}, Optimum Network Simulation and Engineering Tool, http://www.opnet.com

[27] V. Erceg, L. Schumacher, P. Kyritsi, A. Molisch, D. Baum, A. Gorokhov, C. Oestges, Q. Li, K. Yu, N. Tal, B. Dijkstra, A. Jagannatham, C. Lanzl, V. Rhodes, J. Medbo, D. Michelson, M. Webster, E. Jocobsen, D. Cheung, C. Prettie, M. Ho, S. Howard, B. Bjerke, J. Lung, H. Sampath, S. Catreux, S. Valle, A. Poloni, A. Forenza and R. Heath, "TGn Channel Models," *IEEE 802.11 - 03/940r4*, May 2004.