Quality of Service Control in Wireless Local Area Networks

Master Thesis

Author:
Xiang Fan

Supervisors:
Drs. F. Roijers (TNO Telecom, first supervisor)
Prof. dr. ir. B.R.H.M. Haverkort (UT)
Prof. dr. J.L. van den Berg (UT, TNO Telecom)
Dr. ir. G.J. Heijenk (UT)

Design and Analysis of Communication Systems
Faculty of Electrical Engineering, Mathematics and Computer Science
University of Twente

November 2004, Enschede
Abstract

Today wireless internet access, especially the IEEE 802.11 family WLANs, has become widely available. Public hotspots emerge everywhere, from hotels to airports. In addition, private usage also grows rapidly due to the offered convenience.

The current WLAN standard 802.11b does not offer service differentiation. Therefore the quality of an offered service is often not guaranteed. However, a new standard 802.11e has been proposed to improve the quality of service provisioning. This new standard introduces different Traffic Classes for different services and can prioritize the traffic accordingly.

This report covers both the 802.11b and the 802.11e WLANs. In the first place it provides investigation on file transfer times with TCP over 802.11b. This includes both modelling and dynamic simulations. These models are shown to give accurate results for large files. The second part focuses on 802.11e. The differentiation mechanism as well as its performance with realistic service settings is examined through simulation studies. The results provide some important characteristics of the 802.11e parameters and advices for tuning an 802.11e network.
Preface

This thesis is the result of a final project for the department of Electrical Engineering, Mathematics and Computer Science at the University of Twente (UT). The work is carried out at TNO Telecom in Delft, the Netherlands. The research is a part of the Beyond 3G project, which is participated by both UT and TNO Telecom among others.

It is an interesting time at TNO Telecom. Not only have I advanced in the scientific field, I have also learned a lot with regard to working in a professional organization. However, I could not have done it alone without the help and support of many people.

First of all, I would like to thank my supervisors: Prof. dr. ir. Boudewijn Haverkort, Prof. dr. Hans van den Berg, Dr. ir. Geert Heijenk and Drs. Frank Roijers.

During my stay at TNO Telecom, Frank has fulfilled the role as my main supervisor. Your endless patience and in-depth knowledge have guided me through many obstacles alone the way to the end of the research. Our discussions always lead to some great follow-up ideas. I don’t know where this report would have been without your support. Hans has also been very helpful at TNO Telecom. Your advices continually remind me of the larger framework of the project so I do not lose the overview. Geert, your advices on my report are greatly appreciated. I will also not forget your help on both the project planning and scientific input.

Besides the scientific guidance, I have received great support at TNO with regard to personal development. I have learned so much from the personal coaching by Nico Zornig. I also thank the IVG team for giving me the unique opportunity to have a business trip to China. Ian, Michael, I will never forget the days we spent there.

This work could not have been completed without the pleasant working environment. Special thanks go to my roommates, Dany, Joost, Wemke, Anja. There are always enough laughs in our room for a comfortable atmosphere. I am also glad that the contact does not only stay in the office hours. Many members of the KIT-NAK group deserve my gratefulness too, for your plenty of useful advices and fresh ideas.

At last but not the least, I would also like to thank my family during the conduction of this project. Especially my mother, thanks for the endless help in almost everything. And of course Audrey, you are always there to motivate me whenever I need it. Thank you all!
Abbreviations

ACK  ACKnowledgement (packet)
AIFS  Arbitrary IFS
AP    Access Point
BE    Best Effort
BSS   Basic Service Set
CAP   Controlled Access Period
CBR   Constant Bit Rate
CCK   Complementary Code Keying
CFP   Contention Free Period
CI    Confidence Interval
CP    Contention Period
CSI   Carrier Sense Multiple Access with Collision Avoidance
CTS   Clear To Send
CW    Contention Window
DCF   Distributed Coordination Function
DIFS  Distributed IFS
DSSS  Direct Sequence Spread Spectrum
EDCF  Enhanced Distribution Coordination Function
EDCA  Enhanced Distribution Channel Access
EIFS  Extended IFS
ESS   Extended Service Set
FHSS  Frequency Hopping Spread Spectrum
FTP   File Transfer Protocol
GPS   Generalized PS
HC    Hybrid Coordinator
HCCA  HCF Controlled Channel Access
HCF   Hybrid Coordination Function
IEEE  Institute of Electrical and Electronics Engineers
IETF  Internet Engineering Task Force
IFS   Interframe Spacing
IP    Internet Protocol
ITU-T International Telecommunication Union, Telecom standardization
LAN   Local Area Network
LLC   Link Logic Control
MAC   Medium Access Control
NAV   Network Allocation Vector
OSI   Open Systems Interconnection
PC    Point Coordinator
PCF   Point Coordination Function
PIFS  PCF IFS
PLR   Packet Loss Rate
PS    Processor Sharing
QoS   Quality of Service
QSTA  QoS STAtion
RFC   Request For Comments
RTS   Request To Send
RTT   Round Trip Time
SIFS  Short Inter Frame Spacing
STA   (Wireless) Station
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TC</td>
<td>Traffic Class</td>
</tr>
<tr>
<td>TCP</td>
<td>Transmission Control Protocol</td>
</tr>
<tr>
<td>TXOP</td>
<td>Transmission Opportunity</td>
</tr>
<tr>
<td>UDP</td>
<td>User Datagram Protocol</td>
</tr>
<tr>
<td>VCH</td>
<td>Virtual Collision Handler</td>
</tr>
<tr>
<td>VoD</td>
<td>Video on Demand</td>
</tr>
<tr>
<td>VoIP</td>
<td>Voice over IP</td>
</tr>
<tr>
<td>WLAN</td>
<td>Wireless LAN</td>
</tr>
<tr>
<td>WM</td>
<td>Wireless Medium</td>
</tr>
</tbody>
</table>
Contents

1 Introduction .......................................................................................................................... 11
  1.1 Scope ............................................................................................................................. 11
  1.2 Research objective ........................................................................................................ 12
  1.3 Related work ................................................................................................................ 12
  1.4 Overview of the report ................................................................................................ 12

2 Preliminaries ....................................................................................................................... 13
  2.1 Services ......................................................................................................................... 14
  2.2 Quality of Service ....................................................................................................... 14
  2.3 Transport protocols UDP and TCP .............................................................................. 16
      2.3.1 User Datagram Protocol (UDP) ......................................................................... 16
      2.3.2 Transmission Control Protocol (TCP) ................................................................. 17
  2.4 IEEE 802.11 Wireless Local Area Networks ............................................................. 19
      2.4.1 IEEE 802.11 standard, 1999 version .................................................................... 20
      2.4.2 IEEE 802.11b ................................................................................................. 25
      2.4.3 IEEE 802.11e ................................................................................................. 26
      2.4.4 Other 802.11 task groups .................................................................................. 29
  2.5 Network simulator 2 (NS-2) ......................................................................................... 30
      2.5.1 IEEE 802.11 in NS-2 ....................................................................................... 30

3 TCP file transfer times over IEEE 802.11b WLAN .......................................................... 31
  3.1 UDP over WLAN .......................................................................................................... 31
      3.1.1 Basic access mode .............................................................................................. 32
      3.1.2 RTS/CTS mode ................................................................................................. 34
      3.1.3 Summary .......................................................................................................... 35
  3.2 File transfer times with TCP over WLAN ................................................................. 35
      3.2.1 A packet/flow level model approach ................................................................. 36
      3.2.2 TCP packet level studies .................................................................................. 38
      3.2.3 Numerical flow level results of the integrated ordinary PS model ................. 46
      3.2.4 TCP file transfers over WLAN, packet-level revisited .................................... 47
      3.2.5 Refined TCP packet/flow level integrated model ............................................ 49
      3.2.6 Summary .......................................................................................................... 51
  3.3 Conclusions .................................................................................................................. 52

4 Wireless Local Network 802.11e Simulation Studies ....................................................... 53
  4.1 Extensions of NS-2 ....................................................................................................... 53
      4.1.1 Code walk through .............................................................................................. 53
      4.1.2 Running simplified cases ................................................................................... 54
  4.2 Investigation of the impact of the parameters in IEEE 802.11e .................................. 54
      4.2.1 Discussion of the IEEE 802.11e parameters ..................................................... 54
      4.2.2 Scenario description ......................................................................................... 56
      4.2.3 Simulation results .............................................................................................. 57
      4.2.4 Summary .......................................................................................................... 64
  4.3 Performance evaluation of a realistic 802.11e scenario .............................................. 66
      4.3.1 Models .............................................................................................................. 66
      4.3.2 Simulation results .............................................................................................. 67
      4.3.3 Summary .......................................................................................................... 73
  4.4 Conclusion .................................................................................................................... 74
5  Concluding remarks ........................................................................................................ 75
5.1  File transfer times with TCP over IEEE 802.11b WLAN........................................ 75
5.2  IEEE 802.11e WLAN............................................................................................... 75

References.......................................................................................................................... 77
1 Introduction

1.1 Scope

During the years, a wide variety of services are introduced on the Internet, e.g. online shopping, internet banking services, travelling services. In addition, the offering of multimedia services grows rapidly. Not only downloading music or movies, but also streaming services and even real-time conversation are available nowadays on the Internet.

These services have quite diversified characteristics and specific quality requirements, e.g. requirements on throughput, loss, delay and delay variation. Because of the limited resources and the complex nature of the Internet, adequate service provisioning, especially for services with strict Quality of Service (QoS) requirements, is a challenging research topic.

Although only the services are used by the end-users, their performances are affected by the underlying structure of the Internet. Within this scope an important role is played by Transmission Control Protocol (TCP), which is designed for reliably sending data over the Internet. For instance, it is used in applications for browsing the web and downloading files. TCP has a feedback mechanism that is able to detect lost packets and retransmit them. Furthermore, TCP is able to dynamically adapt its packet sending rate to the network conditions, in order to achieve the highest possible throughput. TCP has quite a complex behaviour, especially when it is applied over a medium where TCP’s performance is affected by the medium’s characteristics. The performance of TCP will have a significant influence on the end-user’s perceived quality.

Internet access is initially only available on fixed lines. Recently, wireless medium access is growing explosively for both private and public usage. In particular, IEEE 802.11 Wireless Local Area Network (WLAN) has become increasingly popular with both network operators and end-users. Based on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) technique, WLAN provides a flexible medium for wireless Internet access.

In order to integrate QoS enhancements into the existing 802.11 WLAN standard, IEEE 802.11 Task Group e (IEEE 802.11e) is formed. As opposed to 802.11, 802.11e applies different parameters to different streams to provide prioritized traffic over the WLAN medium, and hence more adequate provision of services with stringent requirements.

In order to offer adequate service provisioning, the telecom operators must set up their networks accordingly. In networks where TCP and WLAN are utilized, it is important to gain insight into their behaviours and their impact on the performances of the offered services. This will be the focus of this research.
1.2 Research objective

There are two objectives of this research:

1. **Performance of TCP over IEEE 802.11b**
   Investigate the performance of services using TCP over IEEE 802.11b WLAN. Relevant performance measures are throughput, fairness and the file transfer time.

2. **Performance of 802.11e, the IEEE WLAN with QoS enhancement**
   Investigate the effects of the parameters defined in IEEE 802.11e, namely $CW_{\text{min},\text{RC}}$, $CW_{\text{max},\text{RC}}$, $AIFS_{\text{RC}}$ and $TXOP_{\text{limit},\text{RC}}$ by means of throughput analysis and examine the performance of 802.11e with a scenario consisting of different types of services.

1.3 Related work

IEEE 802.11b WLANs are extensively studied since its introduction. [Bianchi], [Wu] have proposed to model its access mechanism with a Markovian Chain and it yields accurate results to obtain a WLAN’s effective throughput. Based on this Markovian Chain model, the authors of [Litjens] studied the file transfer times with UDP traffic. Not much is known about the behaviour of TCP flows over WLAN and these studies mostly do not take into the flow dynamics into consideration ([Botti], [Pilosof], [Wu]). The first part of my research will focus on the file transfer times with TCP over WLAN. Similar to [Lassila] and [Haan], where TCP file transfer times are studied over a fixed LAN, flow dynamics will be essential in my study.

The new IEEE 802.11e standard is still in development. Limited studies on its performance are conducted and these studies usually only provide simple simulation results with respect to packet delays ([Gu], [Truong]) and node throughput ([Mangold], [Qiao], [Xiao]). In order to design an optimized 802.11e network, it is important to understand the effects of the tuneable parameters defined in 802.11e. These parameters are extensively examined in my study. Based on the findings, an 802.11e WLAN is designed to satisfy the requirements of a number of services offered over it. Flow dynamics will again be essential in this part of the research.

1.4 Overview of the report

The remainder of the report is organized as follows: section 2 provides relevant background information for our research. The basics of TCP and WLAN will be described. Furthermore, necessary information about services and performance parameters is provided for a better understanding of the other parts of the report. Section 3 presents the results of the performance analysis of TCP file transfers over 802.11b WLAN. The results of an analytical TCP model are discussed. In section 4 we examine the 802.11e WLAN. The impact of the newly defined parameters will be investigated by means of throughput analysis. Subsequently we investigate the performance of a realistic 802.11e scenario by studying performance of a number of pre-defined services. Section 5 concludes this report and provides topics for further research.
2 Preliminaries

This section provides description of services, QoS, transport protocols and WLAN technologies. In order to show the relation between the above topics, the Open System Interconnection (OSI) model can be used ([OSI]). The OSI model defines a networking framework for implementing protocols in seven layers. These layers, together with their relation to the relevant elements, are shown in Figure 2-1.

![Diagram of the OSI stack and its relevant elements with regards to research](image)

**Figure 2-1** the OSI stack and its relevant elements with regards to our research

*WB: Web browsing; VoIP: Voice over IP; VoD: Video on Demand
DL: Data Link layer; LLC: Logic Link Control; MAC: Medium Access Control*

Within the scope of this report, the relevant layers are described in Table 2-1:

<table>
<thead>
<tr>
<th>Layer</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application layer, layer 7</td>
<td>Application and end-user processes are found on this layer. Perceived QoS is identified on this layer.</td>
</tr>
<tr>
<td>Transport layer, layer 4</td>
<td>Provides transparent transfer of data between end systems. TCP is a Transport layer protocol.</td>
</tr>
<tr>
<td>Data Link layer, layer 2</td>
<td>Consists of the MAC and the LLC sublayer. The MAC sublayer controls how a computer on the network gains access to the data and permission to transmit it. The LLC layer controls frame synchronization, flow control and error checking. IEEE 802.11 has a subset of MAC specifications.</td>
</tr>
<tr>
<td>Physical layer, layer 1</td>
<td>This layer conveys the bit stream - electrical impulse, light or radio signal -- through the network at the electrical and mechanical level. It provides the hardware means of sending and receiving data on a carrier. IEEE 802.11 also has a subset of specifications on this layer.</td>
</tr>
</tbody>
</table>

| Table 2-1 The relevant layers in the OSI model |

Section 2.1 provides a short overview of the various types of services, which are offered over the Internet. These services have diversified service requirements and therefore Quality of Service (QoS) measures are necessary for an adequate service provisioning. Within this scope, the service requirements are discussed in more details in section 2.2. In this research the main focus is on the QoS performance issues of TCP and IEEE 802.11e WLAN, thus respectively section 2.3 and 2.4 will present the most important features in these two technologies. In section 2.5 a brief description of the simulation tool used in the research, Network Simulator 2 (NS-2), is provided.
2.1 Services

The variety of services offered over the Internet has rapidly increased during the past few years. The mixture of these services has undergone a clear shift. Besides the conventional data traffic, an increasing amount of multimedia services has emerged on the Internet.

Initially applications over the Internet mainly consist of data transfers, i.e. users will retrieve a webpage for viewing or download programs or other relevant data through FTP. For this type of service the data integrity is very important. The data sending rate for these applications is not constant, which depends on the available link capacity.

As more bandwidth becomes available and the technology advances, many multimedia services are being offered today. Two more distinguishing types have emerged: real-time streaming services and real-time interactive services.

One popular streaming service is Video on Demand (VoD), with which a client can request a movie from a server on the Internet and view it during the reception of data. When a streaming service is active, the server will transmit data packets at a certain rate to the client, usually at a constant bit rate (CBR). However, the link condition between the server and the client varies with the time. The result is that the packets arrive with a variable delay time and the receiving rate is not the same as the sending rate. To ensure that the video is played smoothly, there is often a so-called jitter buffer present at the client to store temporary streaming data packets. The movie is started at the client if the buffer is filled to a certain level. In case that the data packets arrive late due to temporarily congested links, the client is still able to play from the buffer. The buffer is refilled again when link congestion is over.

In interactive services one popular example is Voice over IP (VoIP). VoIP provides telephone services, where users can talk with each other in real-time. Voice signals are processed by an encoder at the source, transmitted through the network and reproduced by a decoder again at the destination\(^1\). Due to the nature of conversations, this type of services has bi-directional data traffic. Compared to streaming services, the streams found in interactive services also have constant bit rate, but usually much smaller packet size.

In our research we focus on the applications web browsing, FTP data retrieval, Video on Demand and Voice over IP.

2.2 Quality of Service

Because of the increasing variety of services offered over the Internet, Quality of Service (QoS) is a challenging topic. According to ITU-T, QoS is defined as stated below (\([\text{ITU-E800}]\)):

“Quality of Service is the collective effect of service performances that determine the degree of satisfaction of a user of the service”

\(^1\) The coder and decoder pair is usually referred as CODEC.
Different types of services have different requirements on the network. It is of great importance for telecom engineers to design a network that can meet these requirements adequately.

If we consider data traffic, there are no specific requirements, besides that the integrity of the files or web pages has to be assured. However, to meet a user’s desired Quality of Service, it is important to have the data transmitted within a certain time boundary, i.e. a minimum *throughput* has to be available for these uses. The *file transfer time* is an important performance indicator for this type of services.

The real-time services have more stringent requirements. In streaming services there is no interaction required between the server and the client. This means that delay is not an important issue in the provisioning of this type of services. However, due to the sensitivity of human ears and eyes the requirements on information loss is very stringent. Loss of information causes distortion in the speech or video signal and the human ears and eyes are only tolerant to a certain amount of distortion. Within a packet switched network like the current internet, information loss is mainly caused by packet loss and therefore the performance measure is defined in terms of *packet loss ratio* (PLR). Besides signal distortion, many CODECs also only tolerate a maximum amount of packet loss.

In interactive services, the quality of a voice conversation is greatly affected by one-way *delay*. If the delay exceeds a certain threshold, it will affect conversational dynamics. The threshold is of the order of several hundreds of milliseconds ([Schneiderman]). In addition, the human ear is very sensitive to short-term delay variation, also known as *jitter*, which is defined as the standard deviation of the delay of packets. Besides delay and jitter, the impact of information loss must also be considered, in terms of the *packet loss ratio*. Packet loss is related to delay and jitter, as packets arrived beyond a certain delay threshold are also considered lost.

Considering these requirements, the ITU-T has provided an indication of suitable performance targets for different types of services. A part of the performance targets is shown in Table 2-2 ([ITU-G1010]):

---

2 Sometimes it is also defined as variance or the range of packet delay times.
### Data services

<table>
<thead>
<tr>
<th>Application</th>
<th>Degree of symmetry</th>
<th>Typical data amount</th>
<th>Key performance parameters and target values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1-way delay</td>
</tr>
<tr>
<td>Web browsing</td>
<td>One-way</td>
<td>~ 15 KB</td>
<td>Preferred &lt;2 s</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Acceptable &lt;4s</td>
</tr>
<tr>
<td>Interactive games</td>
<td>Two-way</td>
<td>&lt; 1 KB</td>
<td>&lt; 200 ms</td>
</tr>
<tr>
<td>Bulk data retrieval</td>
<td>one-way</td>
<td>10 KB-10 MB</td>
<td>Preferred &lt;15 s</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Acceptable &lt;60 s</td>
</tr>
</tbody>
</table>

### Streaming services

<table>
<thead>
<tr>
<th>Application</th>
<th>Degree of symmetry</th>
<th>Typical data rates</th>
<th>Key performance parameters and target values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio streaming</td>
<td>One-way</td>
<td>128 Kbps</td>
<td>&lt; 10 s</td>
</tr>
<tr>
<td>Video on demand</td>
<td>One-way</td>
<td>480 Kbps</td>
<td>&lt; 10 s</td>
</tr>
</tbody>
</table>

### Interactive services

<table>
<thead>
<tr>
<th>Application</th>
<th>Degree of symmetry</th>
<th>Typical data rates</th>
<th>Key performance parameters and target values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice over IP</td>
<td>Two-way</td>
<td>64 Kbps</td>
<td>&lt; 150 ms preferred</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt; 400 ms limit</td>
</tr>
<tr>
<td>Video phone</td>
<td>Two-way</td>
<td>384 Kbps</td>
<td>&lt; 150 ms preferred</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt; 400 ms limit</td>
</tr>
</tbody>
</table>

Table 2-2 Performance targets for data, streaming and interactive services (2001).

2.3 Transport protocols UDP and TCP

Transport protocols have an important function in data transmission over the Internet. The protocols, which are mainly used, are User Datagram Protocol (UDP) and Transmission Control Protocol (TCP). We will explain these two protocols in more details below.

2.3.1 User Datagram Protocol (UDP)

User Datagram Protocol is a connectionless protocol. UDP is developed to work with IP as the underlying protocol. There is a minimum set of protocol information added by UDP and it does not give any guarantee on delivery or duplicate protection. Within UDP there is no feedback mechanism present. This implies that a UDP source cannot detect the network conditions with regard to congestion or link failure. The sending rate of a UDP stream is determined by the rate at which the application generates data, the capabilities of the source (CPU, clock rate, etc.) and the access bandwidth of the source to the Internet. However, the receiving host does not necessarily receive all the data - when the network is congested, UDP-transmitted data could be lost due to router buffer overflow. Thus, the receive rate is limited by network congestion even if the sending rate is not constrained. UDP is more suited for time stringent services, where

---

3 Data rate depends on the codec types. In certain applications other data rates might be found.

4 Depending on the usage of loss concealment algorithms this value may vary.
retransmissions are pointless due to delay requirements. Hence, in the most real-time interactive services and real-time streaming services UDP is applied. Compared to TCP the overhead added by UDP is very low, only 8 bytes. ([UDP])

2.3.2 Transmission Control Protocol (TCP)

Another vastly applied transport protocol is Transmission Control Protocol (TCP). In contrary to UDP, TCP is a connection-oriented protocol and designed to provide delivery guarantee and duplicate protection. TCP necessitates a connection to be set up and kept alive between two hosts for sending a certain amount of data. With the built-in feedback mechanism, a TCP source can detect if a packet has been lost and retransmit it. The feedback loop is realized by the so-called acknowledgement (ACK) packets sent backwards by the destination station. The feedback mechanism also makes it possible for TCP to detect the network conditions, upon which its sending rate is calculated. TCP adds more overhead to the network than UDP. Besides the additional TCP ACK packets, TCP’s header is also larger than UDP. Depending on the implementation, TCP’s header size varies between 20 and 40 bytes. ([TCP])

The time interval between the moment a signal is sent and the moment a response is received is defined as the Round-Trip Time (RTT). RTT contains two major components: a fixed part that is caused by the propagation delay, and a variable part that is caused by queueing delays, which depend on the degree of congestion at the transmission links. TCP updates the RTT value based on the time difference between the sender to transmit a data packet and receive the corresponding TCP ACK.

A sending station is allowed to transmit more packets during one RTT. The maximum number of outstanding packets, which have been sent and are still not acknowledged, is determined by the window size. After a connection is established, the sender will start to transmit data with a default window size. The window is maintained at the sender and adjusted at the arrivals of TCP acknowledgement packets. This allows TCP to tune the window size dynamically according to the network conditions. The algorithms applied for this process have evolved during the years and hence different in all TCP versions.

TCP Tahoe

Among all the TCP versions applied today, TCP Tahoe can be found in a vast amount of workstations. Our research is also based on this version of TCP. TCP Tahoe made its first appearance in 1988. It added a number of new algorithms and refinements to early implementations. The major improvements include Slow Start, Congestion Avoidance and Fast Retransmit. There is also a refinement in RTT estimation included that improves TCP performance. Below we will give a brief description of these features.

Slow start

As stated before, the sender in each TCP connection maintains a window of the maximum number of outstanding packets. This is called the congestion window (cwnd). At the same time there is also an advertised window by the receiver. When a new TCP connection is established in a network, the cwnd is initialized at one packet. Each time a TCP ACK is received by the sender, it increases its cwnd by one. This provides an exponential increase of cwnd during slow start in each RTT. The maximum number of ongoing packets is increased accordingly. The congestion window is said to

---

To avoid confusion with the acknowledgement packets sent on the WLAN MAC, after this section we will refer them respectively by TCP ACK and MAC ACK.

Actually the cwnd is maintained in bytes, but it is always incremented with TCP’s packet size.
be the congestion control imposed by the sender while the advertised window is
congestion control imposed by the receiver.

**Congestion Avoidance**

At a certain point the capacity of the network can be reached (assuming that the
advertised window of the receiver is not a bottle neck here) and the link will begin to
drop some ongoing packets. There are two indications for the sender that packets are
dropped by the link:

1. When a time-out occurs (there is a time-out counter at the sender for each
   packet), or
2. When duplicated TCP ACK packets are received.

If the sender detects that the link is “congested”, TCP’s Congestion Avoidance
algorithm will apply. We first introduce another important parameter of TCP, the *Slow
Start Threshold* \( ssthresh \). At the establishment of a TCP connection \( ssthresh \) is set to a
default value, which is dependent on the TCP implementation. At the moment that the
sender is aware of link congestion, i.e. time-outs or duplicated ACKs have occurred, it
sets \( ssthresh \) to one-half of the current \( cwnd \) and set \( cwnd \) to one. The latter normally
means that the TCP source has to do slow start again \( (cwnd < ssthresh) \). At a given
moment \( cwnd \) will be greater than \( ssthresh \) (half way to the previous congestion point),
and then collision avoidance will take over. From now on \( cwnd \) will increase by \( 1/cwnd \)
at each successful reception of ACK. This results in an increase of one \( cwnd \) in each
RTT, as opposed to the exponential increase during slow start.

In Figure 2-2 the slow start and collision avoidance mechanism of a TCP connection is
shown.

![Figure 2-2 TCP: Slow start and congestion avoidance](image)

**Delayed Acknowledgement**

Normally the TCP receiver does not immediately send an ACK after a TCP data packet
arrives successfully. Instead, it will wait to see if there is data to be sent to the sender
such that the ACK can be sent along. Hence, a TCP ACK will only be generated and
sent by the receiver when:

1. No ACK has been generated for the previously received packet.
2. A certain time interval has expired since the last not acknowledged packet,
   usually the interval is 200 milliseconds.

Basically it states that a TCP ACK packet will be generated for every other received
data packet on a certain connection, unless the delayed timer (usually 200 ms) has
expired. This is called *delayed acknowledgement*. One exception to this algorithm is in
the situation in which an arrived data packet fails the checksum check. In that case a
duplicate ACK is generated and transmitted immediately. In this ACK packet the sender
can find the information of the sequence number of the faulty packet and retransmit it
accordingly.
Fast Retransmit
A duplicate ACK can also be caused by reordering of segments, i.e. due to different routes TCP packets might arrive in another order. Hence, current TCP will not assume packets to be lost when one or two duplicate ACK packets are received. However, if three or more duplicate ACK packets arrive in a row, it will be a strong indication that a TCP data packet is lost. In this case TCP Tahoe will retransmit what appears to be missing, without waiting for the retransmission timer to expire. This is called Fast Retransmit.

Other TCP versions
- TCP Reno
TCP Reno is applied in 1990 first. It retains all enhancements in TCP Tahoe, only the fast retransmit operation is altered to include fast recovery. In this algorithm cwnd will not be decreased to one, but to sssthresh at the occurrence of Fast Retransmit. ([Stevens])

- TCP Vegas
Within TCP Vegas changes have only been made to the sender. After its introduction in 1994, it has been said that TCP Vegas can improve the throughput by 40%-70% and reduce the losses by 50%-80% compared to TCP Reno. The main improvement in TCP Vegas is to give better prediction on timeouts and schedule retransmission more efficiently. One of the major advantages in Vegas is the ability to anticipate congestion, and adjust its transmission rate accordingly ([Vegas]).

2.4 IEEE 802.11 Wireless Local Area Networks
Wireless Local Area Networks have experienced a rapid growth in recent years. Both in private domain and in public hotspots WLANs are used to provide consumers with cheap and fast wireless solutions. The vast majority of the WLANs used today are deployed according to the IEEE 802.11 standard. It is believed that WLAN will keep on growing in the coming years due to its flexibility and economical advantages.

The IEEE 802.11 specification describes two network topologies: ad-hoc and infrastructure mode. A WLAN is operating in ad-hoc mode, if the wireless terminals directly communicate with each other in an independent basic service set (IBSS) without connectivity to the wired backbone network. In infrastructure mode, the wireless terminals communicate via the wired network through an access point (AP). The wireless stations and the AP form a basic service set (BSS) and data transmissions between the stations are via the AP. Sometimes multiple AP’s are connected to each other to form an extended service set (ESS) ([802.11]). Although the AP has more data to transmit than any other wireless station, it does not have a different medium access control than other stations. Infrastructure mode and ad-hoc mode are illustrated in Figure 2-3.
IEEE 802.11 WLAN specifies both a physical layer and a MAC sublayer. During the past years different task groups have been formed to extend it for different purposes. The enhancements from the main task groups are shown in Figure 2-4.

2.4.1 IEEE 802.11 standard, 1999 version

In the original 802.11 specification a WLAN has an access rate of 1Mbps or 2Mbps. For the WLAN physical layer, three techniques are specified: Direct Sequence Spread Spectrum (DSSS), Frequency Hopping Spread Spectrum (FHSS) and infra-red (IR). ([802.11]). In the 802.11 standard two MAC layer protocols are defined, namely the Distributed Coordination Function (DCF) and the optional Point Coordination Function (PCF). The latter is hardly used in current products.

- Distributed Coordination Function
  DCF is designed to support asynchronous data on a best effort basis. Today most of the IEEE 802.11 devices operate in this mode only. DCF is based on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol.

  The most important part of this procedure is the backoff mechanism applied before the transmissions of packets. When a station (STA) has a packet to send, it will first listen to the medium. If the medium is busy, it will start a defer backoff procedure. After detecting that the wireless medium has been idle for a consecutive DCF Interframe Spacing (DIFS) time, the STA keeps sensing the medium for an additional random time called the backoff period. The packet will be transmitted if the medium remains idle after the backoff period is expired.
The length of the backoff period is determined by the so-called Contention Window (CW), which is maintained separately at each wireless station in a BSS. A random integer value (backoff window) is drawn uniformly from the interval [0, CW], and this is the number of timeslots in a backoff period. If the medium is sensed idle for the duration of a timeslot, the backoff window will decrement by one and this process continues as long as the medium remains idle. When the backoff window reaches 0, the station will transmit the data packet. If the wireless medium is sensed busy during the backoff process, the count down of the backoff window will be suspended. It resumes with the residual value when the medium is sensed idle for a consecutive DIFS. Occasionally stations have to wait for an Extended Interframe Spacing (EIFS) instead of a DIFS before decrementing the (residual) backoff window. This takes place after the station detected that a frame originated from other stations was not received correctly. EIFS has a much larger duration than DIFS.

802.11 specifies that, if a packet arrives while the medium is idle, the station will only need to wait for a DIFS of idle time before starting transmission. No backoff window will be drawn in this case.

The initial CW value (\( CW_{\text{min}} \)) is identical for all stations in an IEEE 802.11 WLAN. Hence all stations will have equal probability to access the medium and thus equal number of transmitted packets on the long turn.

There is a probability that multiple sources transmit packets at the same time. While a STA is transmitting, it can not sense the activity of other stations due to interference with its own signal. Hence it cannot detect collision during transmission. After receiving a packet successfully, the destination station waits a short interframe spacing (SIFS) time and sends an acknowledgement (MAC ACK) packet to the sender. A SIFS is shorter than a DIFS such that the transmission of the MAC ACK packet will not be interrupted by other data transmissions. This means that if the source station does not receive a MAC ACK packet in time, it can conclude that the packet did not arrive correctly. In this case it will calculate a new CW with:

\[
CW_{\text{new}} = \min\{CW_{\text{max}}, 2 \times (CW_{\text{old}} + 1) - 1\},
\]

in which \( CW_{\text{max}} \) is the upper boundary for the exponential growth of CW. The station starts a new backoff procedure with \( CW_{\text{new}} \). IEEE 802.11 specifies a maximum number of retry attempts (retry counter) for transmitting each packet, after which the packet will be dropped. The number of retry attempts depends on the packet size.

The backoff process is illustrated in Figure 2-5, where several packets are sent by the stations towards the AP.
Figure 2-5 The IEEE 802.11 DCF backoff process in basic mode

The operation mode described above is called Basic Access mode. Furthermore, there is another mode defined in the IEEE 802.11 specification, namely the RTS/CTS mode. Request To Send (RTS) and Clear To Send (CTS) are additional control frames in this scheme compared to basic access mode. A STA can reserve medium bandwidth prior to a transmission of a datagram by utilizing these control frames. When operating in RTS/CTS mode, a STA sends an RTS packet first after successfully contending for the wireless medium. The RTS packet carries information of the transmission duration of the awaiting data packet and the destination station. All other stations in the same BSS will read this information from the RTS packet and update their Network Allocation Vector (NAV) accordingly. In its NAV a STA will store the information how long the medium will be occupied. After a SIFS, the destination STA send a CTS packet to the sending station. Again, there is a duration field in the CTS packet header and this information will be used by all stations in the BSS to update their NAV. If the CTS packet arrives successfully at the source station, it can then transmit its data packet without any interruption by other stations. The RTS/CTS scheme is also illustrated in Figure 2-6, where several packets are transmitted by the stations towards the AP.

RTS/CTS modes can prevent the hidden STA and exposed STA problems, see Figure 2-7 and Figure 2-8.
In basic access mode C cannot detect that A is transmitting towards B. Therefore it will assume the medium is free and transmit its own packet towards B and a collision will take place at B.

A big disadvantage of basic mode is that a collision will take up the transmission time of a whole data packet. In a network where many stations are contending for the wireless medium, collisions will take place frequently. The loss on collisions could be very significant.

When operating in RTS/CTS mode, a collision will only involve an RTS packet. Due to its small size (RTS is 20 bytes), the waste of bandwidth is much more limited than a collision on a full length MAC data packet. RTS/CTS mode is not always more efficient. The additional control frames introduce extra overhead in data transmission. Hence in a network with few contending stations, where collisions do not frequently take place, the RTS/CTS scheme will have a negative impact on the medium efficiency. Payload size of the data packets should also be taken consideration, since the overhead introduced by the RTS/CTS packets will not compensate the waste of collision on small data packets. Therefore the IEEE 802.11 specification defines a manageable threshold, such that larger packets are transmitted in RTS/CTS mode and smaller packets in the basic access mode.

Within the DCF, there is no mechanism to guarantee minimum delay to support stations where time-bounded services are desired.
• Point Coordination Function

Another coordination function defined in 802.11 is the Point Coordination Function (PCF). PCF is based on poll-and-response mechanism. In each BSS there is a point coordinator (PC). Usually the function of the PC is performed by the access point in the BSS. Within the PCF the PC will assign other stations timeslots to transmit on a contention-free (CF) base.

PCF will coexist with DCF and takes place at regular intervals where it is applied. The PCF is used in so-called contention-free period (CFP) alternated by DCF in contention period (CP).

The PC starts a CFP by sending a beacon frame. This is done when the medium is sensed idle for a PCF interframe spacing interval (PIFS). As a PIFS is shorter than DIFS, the PC is able to start a CFP before other stations can start contending for the wireless medium. The stations in the BSS will retrieve the duration information of the coming CFP from the beacon frame. The order of polling is determined by a polling list maintained at the PC. When a SIFS is elapsed after the transmission of the beacon frame, the PC starts CF transmission, by sending a CF-poll (no data), data, or data +CF-poll frame to the first station in the BSS. After receiving one of the mentioned frames, the destination station responds after a SIFS interval by sending a CF-ACK (no data) or a data+CF-ACK frame. When these packets are received successfully by the PC, it can repeat the same procedure to the same station or to the next station in the polling list. If the PC does not receive anything within a PIFS interval from the polled station, it will continue to poll the next station on the polling list. The PC can terminate the CFP by sending a CF-END frame, at any time instance after the medium is detected idle for a SIFS interval ([802.11], [Prasad]). The PCF scheme is illustrated in Figure 2-9.

Because PCF is a central-controlled method, it can offer limited priority to stations where time-stringent services are desired, e.g. by polling such stations more frequently, thus also give them more opportunities to receive and transmit data. However, many inadequacies have been identified, such as shortened CFP, unknown transmission duration of polled stations and difficulties to predict the amount of frames STAs want to send. Even more, there is no management interface defined to build up and control PCF operations. This effectively results in a very limited deployment of PCF in current WLANs.
2.4.2 IEEE 802.11b

In 1999, 802.11 Task Group b was formed to develop WLANs that are capable of operating at higher access rate, up to 11Mbps. 802.11b WLANS provides additional data rates of 5.5 and 11 Mbps, depending on the signal strength in the specific network. The PHY layer is enhanced with High Rate-Direct Sequence Spread Spectrum (HR-DSSS), also known as Complementary Code Keying (CCK). However, the MAC mechanisms described in section 2.4.1 are unchanged for IEEE 802.11b, so generally 802.11b can be treated as legacy 802.11 with higher access rate ([802.11b], [Prasad]). Currently 802.11b WLAN is very widely applied and has replaced the most legacy 802.11 WLANS.

Not all bits sent by a station can profit from the higher access rates. The 802.11 PHY defines a packet structure with three major parts:
1. **Preamble**, serves for signal detection and synchronization;
2. **Header**, carries information about data-rate and packet;
3. **Payload**, stores the real data.

Two different preambles and headers are defined: the mandatory Long Preamble and header and an optional Short Preamble and header. In the case of Long Preamble and header a long preamble of 144 bits will be used. Both preamble and header are transmitted with 1 Mbps data rate. If Short Preamble and header is used, the preamble only contains 72 bits and is sent at 1 Mbps, the header of the packet will be transmitted at 2 Mbps. In both cases only the payload of the packet is transmitted at the higher data rate. See Figure 2-10 and Figure 2-11:

In legacy 802.11 WLANS only Long Preamble and header can be found. Hence the most 802.11b stations mainly support this mode to assure compatibility with legacy 802.11 WLANS. The Short Preamble and header mode is intended for operations where maximum throughput between the stations is desired and interoperability with legacy equipments is not a consideration. Our research is mainly based on the 802.11b standards and only the Long Preamble and header mode will be considered. The relevant parameters of 802.11b are summarized in Table 2-3.
### IEEE 802.11e

IEEE 802.11 TGe was formed in 2001. The main purpose of 802.11e is to integrate QoS enhancements and multimedia supports into the existing WLAN standards, i.e. IEEE 802.11b and IEEE 802.11a. QoS and multimedia support are critical to wireless networks, since a mixture of voice, video, audio and data will form the content of traffic over WLAN in such environment. IEEE 802.11e has been through several drafts and is scheduled for approval in 2004.

802.11e defines additional features above specified in the 1999 edition of IEEE 802.11. These enhancements distinguish *QoS enhanced Stations* (QSTAs) from non-QoS STAs (STAs), and *QoS enhanced Access Point* (QAP) from non-QoS access point (AP). These features are collectively termed QoS facility. In IEEE 802.11e a new channel access function is defined for a *QoS aware BSS* (QBSS), namely the *Hybrid Distribution Function* (HCF). HCF is an enhancement on the MAC sublayer. Consequently, 802.11e stations remain fully backward compatible with IEEE 802.11 wireless standards ([IEEE]). Within HCF there are again two operation modes: **Enhanced Distributed Channel Access (EDCA)** is a contention-based channel access function that operates concurrently with **HCF Controlled Channel Access (HCCA)**. The latter is based on a polling mechanism, which is controlled by the *Hybrid Coordinator* (HC). EDCA is sometimes also referred as **Enhanced Distribution Coordination Function (EDCF)**. HCF/EDCA mode is comparable with DCF within the legacy 802.11 standard and HCF/HCCA mode is based on the PCF MAC mechanism.

Before discussing the channel access functions in more details, another new concept in IEEE 802.11e has to be introduced: *Transmission Opportunity* (TXOP). A TXOP is a time interval in which a particular QTA has the right to initiate transmissions onto the wireless medium ([802.11e]). A TXOP is defined with a starting time and a duration length within a QBSS, a QSTA will acquire a TXOP differently under the two HCF modes.

- **HCF/EDCA**

  A TXOP can be gained through a contention mechanism similar to DCF in 802.11. This is regulated by HCF/EDCA within 802.11e. A TXOP acquired in this mode is called an **EDCA-TXOP**. Compared to DCF, HCF/EDCA provides prioritized QoS,

---

<table>
<thead>
<tr>
<th>Parameter</th>
<th>value</th>
<th>Parameter</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Rate</td>
<td>{1, 2, 5.5, 11} Mbps</td>
<td>Retry limit(^7)</td>
<td>{4, 7}</td>
</tr>
<tr>
<td>Basic rate(^8)</td>
<td>{1, 2} Mbps</td>
<td>PHY header</td>
<td>48 bits</td>
</tr>
<tr>
<td>Time slot</td>
<td>20 (\mu)s</td>
<td>PLCP preamble</td>
<td>{72, 144} bits</td>
</tr>
<tr>
<td>SIFS</td>
<td>10 (\mu)s</td>
<td>MAC header</td>
<td>{224, 272} bits</td>
</tr>
<tr>
<td>PIFS</td>
<td>30 (\mu)s</td>
<td>RTSThreshold</td>
<td>2400 bits typical</td>
</tr>
<tr>
<td>DIFS</td>
<td>50 (\mu)s</td>
<td>MAC ACK</td>
<td>112 bits</td>
</tr>
<tr>
<td>EIFS</td>
<td>304 (\mu)s</td>
<td>RTS</td>
<td>160 bits</td>
</tr>
<tr>
<td>CWmin</td>
<td>31</td>
<td>CTS</td>
<td>112 bits</td>
</tr>
<tr>
<td>CWmax</td>
<td>1023</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^7\) Depends on data frame size. Large frames have a retry limit of 4 and small frames 7.

\(^8\) This refers to the data rate at which the PHY headers are transmitted.
i.e. QoS based on priority of access to the wireless medium. Prioritized QoS is offered by introducing Traffic Classes (TC). Each TC will have an independent queue in the station and its own set of parameters. By setting different values to these parameters in different TCs, 802.11e will assign the TCs with different priority.

Upon packet arrival at a certain QSTA transmission queue, the QSTA will determine the priority of this packet and place it in the appropriate TC queue. Contentions will take place inside the QSTA between the different TC queues.

In the 802.11e draft it is stated that a maximum of 8 simultaneous TC queues can be maintained within a QSTA. This process is illustrated in Figure 2-12 (there are respectively six and three TCs in station 1 and 2, with TC0 having the highest priority):

The internal contention follows similar mechanism as DCF, which is described in section 2.4.1, only the differentiated values of the queue parameters provide some packets a higher probability to be sent than others. If a collision takes place at this stage, the packet from the queue of the highest priority will be transmitted. The winning station of the contention on the wireless medium will be granted a EDCA-TXOP, in which this station can transmit packets within its $\text{TCTXOP}_{\text{Limit}}$. Collisions at the internal contentions are called virtual collisions. In such cases the CW will be increased (if the maximum is not yet reached) just like the DCF. However, this will not count as a transmission attempt. There is still a probability that a collision takes place on the wireless medium, in this case both CW and retry count will be increased accordingly.

The parameter set contains the following elements$^9$:

1. Arbitary Interframe Spacing Number (AIFSN). The time interval AIFS, defined as 

   $$\text{AIFS}_{\text{TC}} = \text{SIFS} + \text{AIFSN}_{\text{TC}} \times \text{TimeSlot} \text{, with } \text{AIFSN}_{\text{TC}} \in \{1, 2, 3, \ldots\}$$

---

$^9$ In the early versions of 802.11e drafts another tunable parameter is defined: Persistent Factor ($\text{PF}_{\text{TC}}$). This parameter gives a measure how fast CW grows after successive collisions. In legacy 802.11 WLANs we will find a PF of 2. This parameter is dropped out in later draft versions and will not be considered in this report.

$^{10}$ In some papers and drafts the relation between AIFSN and AIFS is represented by $$\text{AIFS}_{\text{TC}} = \text{SIFS} + \text{AIFSN}_{\text{TC}} \times \text{TimeSlot} \text{, with } \text{AIFSN}_{\text{TC}} \in \{0, 1, 2, 3, \ldots\}$$ . In this report we will use the relation given in the body text.
is the minimum time interval between when the wireless medium becomes idle and the start of transmission of the related TC. Similar to DIFS, a station must wait for the medium to be idle for an AIFS before counting down in its (residual) CW.

2. \( CW_{\min} \): Defines the initial interval from which a CW is drawn for the backoff mechanism.

3. \( CW_{\max} \): Defines the maximum interval in the exponential backoff mechanism.

4. \( TXOPLimit \): Defines the maximum duration for which a QSTA can transmit after obtaining a TXOP. The minimum value of TXOPLimit is 0, with which one packet can be transmitted in an acquired TXOP.

These parameters are illustrated in Figure 2-13.

- **HCF/HCCA**

Similar to PCF, HCF/HCCA is a polling-based access mechanism. The hybrid coordinator (usually integrated in the QAC) is able to start a Controlled Access Period (CAP) whenever it is considered necessary. A CAP is comparable with a CFP in the 802.11 standard. According to the draft standard it can even be done within an active HCF/EDCA period, in order to poll traffic with specific QoS requirements. The polling sequence within the HC is based on the information it has received from the QSTAs. The QoS requirements of certain traffic streams within a QSTA are stored in the so-called Traffic Specification (TSPEC) device. A QSTA can initiate a TSPEC negotiation with the HC to pass the information, and the HC on its turn can arrange its polling schedule based on all the TSPEC information it has received.

Just like the PCF, there are still many unsolved issues for the implementation of the HCF/HCCA mode, such as how the HC should manage the polling of a large number of interactive streams without harming applications using EDCA scheme. Furthermore, some real-time services require periodical polling, this could be a
problem if there are many of these services present, and the polling period duration becomes longer than the required inter-poll time. Due to such uncertainties the first generation IEEE 802.11e products are expected to include only the additional HCF/EDCA scheme. Hence, HCF/HCCA will not be considered in this report.

2.4.4 Other 802.11 task groups

- IEEE.802.11 TGa
  This extension defines a PHY layer that operates in the 5 GHz band. However, the MAC mechanism is not modified in 802.11a compared to legacy 802.11\textsuperscript{11}. Commonly, 802.11a WLANs are found to operate at 6, 12 and 24 Mbps although it is capable of operating at 54 Mbps. Compared to 802.11b PHY, \textit{Orthogonal Frequency Division Multiplexing} (OFDM) encoding scheme is used rather than DSSS. Hence, 802.11a equipments are not compatible with legacy 802.11 or 802.11b products. ([Prasad], [IEEE])

- IEEE 802.11 TGg
  An 802.11g WLAN offers similar access rates as 802.11a in the 2.4 GHz band. The main coding scheme used in 802.11g is again same to 802.11a, OFDM. However, to ensure backward compatibility with 802.11b products, 802.11g also supports CCK. The 802.11g supplement to the legacy 802.11 standard is approved in June, 2003.

- IEEE 802.11 TGi
  Security has been a primary concern for many companies to deploy wireless networks. The main focus of task group i is to add more security into 802.11 framework. Enhancements are made on the 802.11 MAC sublayer. The enhancements include authentication, cipher flexibility and scalability. The i amendment to IEEE 802.11 standard is recently approved (June, 2004) and will be published soon. ([IEEE], [802.11i], [Prasad])

- IEEE 802.11 TGn
  This task group is formed to achieve even higher access rate while preserving the backward interoperability with legacy 802.11 and 802.11 a/g products. The goal is to bring the maximum data rate at the MAC sublayer up to at least 100 Mbps. Current 802.11x standards are all specified with an access rate measured at the PHY layer. The achievable throughput at the MAC layer is far less. For example, the 802.11b standard, with a physical-layer rate of 11 Mbps, typically delivers a net maximum 5 to 6 Mbps, while the 802.11a and 802.11g standards, with a peak PHY data rate of 54 Mbps, deliver a maximum throughput of around 20 to 24 Mbps. TG n is still active and expects to have an approved standard by the end of 2005 ([802.11n]).

\textsuperscript{11} Due to the different physical layer implementation, timeslot, SIFS etc. all have different values in 802.11a than in legacy 802.11 and 802.11b.
2.5 Network simulator 2 (NS-2)

Network Simulator 2 (NS-2) [NS-2] is a simulation tool originated from Lawrence Berkeley National Laboratory. It is targeted at networking research and based on discrete events simulations. NS-2 provides substantial support for simulation of routing and multicast protocols over wired and wireless networks. NS-2 has an advanced TCP module, which is applied and verified extensively in the network community. Because simulation with TCP is essential in our research, NS-2 is an excellent simulation tool within this scope.

NS-2 is a popular tool among the researchers in both the wired and wireless world and has been used to obtain vast amount of simulation results used in various papers. However, NS-2 undergoes changes everyday and is not a finished and polished product. In order to gain insight into NS-2’s reliability in wireless simulations, simulation results are verified with expectations from several known mathematical models. This will be described in more details in section 3.1.

The idea of a discrete event scheduler is that actions may only be started as a result of an event. In NS-2 this is taken care of by a scheduler and a scheduling list. Events are inserted into scheduling list upon request, together with their expiration time. The scheduler is responsible to go through the list and perform the necessary actions.

2.5.1 IEEE 802.11 in NS-2

Within the DCF implementation, a packet is handed by the LLC through an advanced drop-tail interface queue to its MAC. Beacons are not really included but routing update messages are transmitted in each interval of several seconds.

The DCF MAC protocol can handle DATA/ACK/RTS/CTS, as well as the broadcasting type of packets. Both physical and virtual carrier sense are supported. The interframe space intervals SIFS, PIFS, DIFS are implemented accordingly. EIFS is also implemented and applied after each unsuccessful transmission attempt.

Within each station only one queue is present for all the awaiting packets. This queue is controlled by several timers defined in the NS-2 MAC: defer time, backoff timer, interface timer, send timer and NAV timer. These timers start upon certain events and are responsible to insertion of new events into the scheduling list. There is also a procedure present to cancel/pause these timers. After a timer expires, a handling procedure is called to execute the required follow–up actions.

NS-2 includes all three PHY layer specifications: Frequency Hopping Spread Spectrum (FHSS), Direct Sequence Spreading Spectrum (DSSS) and Infra-red (IR). In the DSSS PHY, which is relevant to our research, both short and long preambles are included. Long preamble is default if it nothing is specified.

Within NS-2 a capture threshold is also implemented\(^\text{12}\). It relies on the received power strength and may retrieve one packet correctly even if it collides with other packets. However, this effect can be excluded by applying an extremely high capture threshold.

\(^{12}\) This is different than receiving threshold, above which a packet can be received correctly without collision.
3 TCP file transfer times over IEEE 802.11b WLAN

On the current Internet the major part of traffic is controlled by TCP. TCP also plays an important role in the data flow over WLAN and its performance has a direct effect on an end-user’s perceived quality. Although on wired links TCP has been studied extensively, very few research results are known about its behaviour over IEEE 802.11 WLAN.

 Particularly, the mean transmission delay or from a user’s point of view simply the download time of a data file is an interesting and practical research topic. Similar research is done for UDP over WLAN by the authors of [Litjens]. Due to its congestion control mechanism different results are expected from TCP.

The TCP behaviour is analytically approached with an integrated packet/flow level model. Packet level investigation captures more details of the system, e.g. WLAN and TCP parameters, but does not take into account the flow level dynamics. Thereby a fixed number of persistent flows is assumed. Flow level study considers the flow dynamics related to the arrival and departure of users or flows.

This part is organized as follows. In section 3.1 the behaviour of UDP over WLAN is examined, together with a few models concerning UDP over WLAN. We continue our research with TCP file transfers over WLAN in section 3.2. It provides the findings on the packet level investigation of TCP, as well as for predicting mean file transfer times with TCP flows.

3.1 UDP over WLAN

Recently Bianchi and Wu et al. ([Bianchi], [Wu]) have studied the IEEE 802.11 “saturation throughput”. This is defined as the limit reached by a certain system as the offered load increases, and represents the maximum load the system can carry in stable conditions. Persistent UDP flows are studied in their work. This means that the sources always have packets ready to send in their queue. It is shown that their models give a fairly accurate approximation for the WLAN saturation throughput. We use NS-2 2.27 to determine the saturation throughput and compare it to the analytical results, in order to gain insight into the reliability of NS-2.

The scenario is depicted in Figure 3-1.
We consider a WLAN BSS consisting of different numbers of wireless stations and one access point. In order to achieve comparable results with [Bianchi] and [Wu] the WLAN operates with an access rate of 1 Mbps. Accordingly, the parameters used in the following paragraphs are calculated based on the 1Mbps access speed, unless it is mentioned otherwise.

One of the most important findings from these papers is that the number of sources has a significant impact on the “saturation throughput”. The relation follows approximately a Markovian Chain model and the saturation throughput can be derived for a certain number of present stations in the BSS. Especially in WLAN’s Basic access mode the drop of saturation throughput due to increased number of active transmitting stations cannot be neglected.

In NS-2 similar scenarios can be simulated. We have obtained results for WLAN in both Basic access mode and RTS/CTS mode. The UDP flows are in the upstream direction\(^{13}\). First of all we discuss the results of Basic access mode.

3.1.1 Basic access mode

The scenario is based on the default WLAN provided with the NS-2 package. The relevant parameter values used for our simulations are given in Table 3-1. These parameters are also used in the mathematical model to obtain comparable results.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>value</th>
<th>Parameter</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Rate</td>
<td>1 Mbps</td>
<td>PHY header</td>
<td>48 bits</td>
</tr>
<tr>
<td>Basic rate</td>
<td>1 Mbps</td>
<td>PLCP preamble</td>
<td>144 bits</td>
</tr>
<tr>
<td>SIFS</td>
<td>10 µs</td>
<td>UDP payload</td>
<td>1500 bytes</td>
</tr>
<tr>
<td>DIFS</td>
<td>50 µs</td>
<td>IP header</td>
<td>20 bytes</td>
</tr>
<tr>
<td>EIFS</td>
<td>304 µs</td>
<td>UDP header</td>
<td>8 bytes</td>
</tr>
<tr>
<td>MAC overhead</td>
<td>224 bits(^{14})</td>
<td>IFQ length</td>
<td>50</td>
</tr>
<tr>
<td>Propagation delay</td>
<td>2 µs</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 3-1 Simulation parameters for NS-2 UDP scenario.*

The simulation results, together with the numerical results from Wu’s model, are plotted in Figure 3-2. The results are obtained with ten replications, which result in a confidence interval (95%) to mean fraction of below 1%. In the remaining of section 3 the same fraction ratio is also achieved unless mentioned otherwise. In Figure 3-2 we also show the results of a simulator used in [Litjens]. The explanation follows below.

\(^{13}\) The directions are defined from the user’s point of view. Hence downstream is from the distant web server or access point to the STA, and upstream is in the opposite direction.

\(^{14}\) In 802.11 specification section 7.2.2 states that the MAC header omits field address 4 if transmission only takes place in the same BSS. This means that the overhead from the MAC layer is 28 bytes.
We see that the saturation throughput decreases if the number of contending stations increases. This confirms the findings from [Wu] and [Bianchi]. There are mainly two factors that determine the shape of the curve. Due to WLAN’s contention mechanism, the number of collisions increases among the number of active stations in the BSS. Therefore more bandwidth is wasted on collisions. On the other hand, if there are more stations competing for the medium, they fill in each other’s post backoff and therefore increase the medium efficiency. These two effects have opposite influence on the saturation throughput. It shows clearly that in basic access mode the first effect is dominant. This results in the steadily declined shape of the medium saturation throughput with an increased number of stations. In the same figure the theoretical throughput, calculated from the model of Wu’s, is also plotted. What we see is that for a lower number of stations in the scenario, up to 15 stations, the difference between the model and our simulation is very insignificant. If the number of stations is further increased, the difference between the results from NS and the model from Wu becomes larger. The difference observed here can probably be explained by the presence of EIFS in NS-2 while it is not taken into consideration by [Wu] and [Bianchi]. In [Litjens] there is another simulator written to analyse 802.11 WLAN behaviours. We have used this tool to determine the influence of EIFS. If EIFS is set equal to DIFS, similar results from both the [Litjens] tool and model Wu have been obtained. If EIFS is set to 304 µs, according to the IEEE 802.11 specification, the [Litjens] tool shows a higher saturation throughput and the curve is even closer to the simulation results from NS-2. The occasional routing packets in the NS-2 simulations could be the reason for the little differences observed above 40 stations.

If a single UDP source is sending towards the AP, the throughput can be determined. In Figure 3-3 transmission of a single UDP packet is shown.

If data is transmitted over 802.11, different types of overhead are added. On the WLAN MAC layer an extra MAC ACK segment is also sent whenever a packet is successfully received.
In our simulation UDP segments of 1500 bytes are used. The added overhead from different layers is listed in Table 3-1. The duration for a transmission cycle of one packet is:

\[
T_{cycle} = T_{data} + T_{MacACK} + SIFS + DIFS + T_{BackOff} + 2 * T_{delay}  \\
= \frac{\text{Payload} + \text{UDPheader} + \text{IPheader} + \text{MACheader}}{\text{datarate}} + \frac{\text{Preamble} + \text{PLCP}}{\text{PLCPdatarate}} + \frac{\text{basicrate}}{\text{MacACK}} + \frac{\text{Preamble} + \text{PLCP}}{\text{PLCPdatarate}} + \frac{\text{SIFS} + \text{DIFS} + T_{BackOff} + 2 * T_{delay}}{\text{datarate}}  \\
= \frac{12000 + 64 + 160 + 224}{1000000} + \frac{144 + 48}{1000000} + \frac{112}{1000000} + \frac{144 + 48 + 10 + 50 + 310 + 4}{1000000}  \\
= 13318 \mu s
\]

The average CW is 15.5 timeslots with a slot length of 20 \( \mu s \).

The theoretical saturation throughput with one UDP source is then:

\[
R_{UDP-B} = \frac{\text{Payload}}{T_{cycle}} = \frac{12000}{13318} = 0.901Mbps
\]

Simulations with NS-2 yield a saturation throughput of 0.896 Mbps. The little difference here can be explained by the routing packets sent occasionally, which cause extra overhead in the NS-2 simulations. NS-2 trace files show that these packets are sent from time to time.

### 3.1.2 RTS/CTS mode

![Figure 3-4 Transmission of a UDP packet through WLAN in RTS/CTS mode](image)

Section 2 has described a second transmission mode specified in IEEE 802.11: the RTS/CTS mode. In this mode two extra packets are sent in each transmission cycle, namely RTS and CTS. See Figure 3-4. Simulations with NS-2 have yielded the following saturation throughput curve against the number of active stations in the scenario.

![Figure 3-5 Saturation throughput with upstream UDP traffic, RTS/CTS mode](image)
If there are few stations in the network, the RTS/CTS mechanism shows a lower throughput than in Basic Mode. The reason here is due to the extra overhead introduced by the RTS and CTS packets. The two effects described in the previous section are again present here. However, in RTS/CTS mode collisions only involve RTS packets. They are relatively small in size. Hence the bandwidth wasted on collision is much less severe compared to basic access mode. The result is that the saturation throughput only drops slightly even if the number of stations in the network is high. The drop of saturation throughput with a high number of STAs is much less than in Basic access mode. In Figure 3-5 we can also see the other effect described in 3.1.1, namely that different sources fill in each other’s post backoff window. This explains the increase of the medium efficiency if the number of contending stations rises from one to up to four.

In RTS/CTS mode simulations results with NS-2 yield a slightly lower saturation throughput overall. The difference is probably again caused by the routing packets sent by the stations. These routing packets do not follow the RTS/CTS scheme and the amount of them increases if there are more stations in the scenario.

If only a single UDP source is present, the theoretical throughput can be determined. Compared to the Basic access mode, two more SIFS and the transmission of RTS and CTS have to be added to the transmission cycle in RTS/CTS mode. The total length of a transmission cycle increases to:

\[
T_{\text{cycle}} = T_{\text{data}} + T_{\text{backoff}} + 3 \times \text{SIFS} + T_{\text{RTS}} + T_{\text{CTS}} + \text{DIFS} + \text{SIFS} + 4 \times T_{\text{delay}}
\]

\[
= 12640 + 304 + 30 + 352 + 304 + 50 + 310 + 8
\]

\[
= 13998 \mu s
\]

The throughput with one UDP source is:

\[
R_{\text{UDP},R} = \frac{12000}{13998} = 0.857 \text{Mbps}
\]

Simulations from NS-2 have shown a throughput of 0.853Mbps with one UDP STA.

### 3.1.3 Summary

NS-2 shows very similar results as the expectations from the mathematical models. NS-2 does tend to give a different saturation throughput if the number of active stations is high, especially in Basic access mode. This is caused by the fact that the models have not considered the influence of EIFS. If that is taken into consideration, the difference becomes very insignificant. In realistic scenarios it occurs rarely that more than 15 contending stations are present. In that region NS-2 shows very comparable results as the mathematical models have predicted. Therefore we can conclude that NS-2 is a reliable simulation tool for our further research.

### 3.2 File transfer times with TCP over WLAN

WLAN performance is largely determined by the throughput at its physical layer and MAC layer. A number of papers have studied the throughput performance of WLAN extensively ([Wu], [Bianchi]). We can see that models with accurate predictions can be made for file transfer times through UDP flows over WLAN ([Litjens]). On the other hand, extensive works have also been conducted to model the behaviour of TCP over wired links. However, there is still little known for the performance of TCP file transfers over WLAN in the literature, especially if it operates in infrastructure mode.
3.2.1 A packet/flow level model approach

As mentioned above, the authors of [Litjens] have developed a fairly accurate model for predicting file transfer times with UDP flows over WLAN. In their research they have proposed an integrated packet/flow level modelling approach. On packet level the WLAN DCF performance can be modelled using a discrete time Markov Chain as described in [Wu] and [Bianchi]. With this model the saturation throughput can be accurately calculated for a given number of persistent flows.

On flow level the authors of [Litjens] have considered WLAN as a queueing model with Poisson flow arrivals and a Generalized Processor Sharing (GPS) service discipline. WLAN 802.11 DCF MAC design is to distribute the transmission capacity fairly among the STAs. Hence it matches the condition of the GPS model that jobs get an equal amount of server capacity. The service rates of these flows are dependent on the number of present flows in the system, which can be retrieved from the packet level model mentioned previously. These state-dependent service rates can then be used to utilize the results of the GPS queueing model. The expected number of flows present in the system can subsequently be determined. With Little’s Theorem ([Cooper]) one could then calculate the expected flow transfer time.

Below is a brief description of the formulas used here:

The GPS model, ordinary PS:

Let jobs arrive according to a Poisson process with rate $\lambda$ at a server with capacity $C$. The job sizes are independent and identically distributed with mean $1/\mu < \infty$. Further we denote $r_n$ as the (joint) transmission rate if there are $n$ jobs in the system. All these jobs will get an equal amount of capacity from the server.

In Cohen’s work ([Cohen]) we can find the solution for the GPS model. First define

$$\phi(n) = \begin{cases} \frac{\lambda}{\mu_n}, & \text{for } n \in N, \\ 1, & \text{for } n = 0, \end{cases}$$

and

$$\Psi(n) = \prod_{i=0}^{n-1} \phi(i).$$

Then the distribution of the number of jobs $N$ in the systems is given by

$$P(N = n) = \frac{\Psi(n)}{\sum_{m=0}^{\infty} \Psi(m)}.$$ 

If the distribution of $N$ is known, the mean job number $\bar{N}$ can be calculated. The mean transmission delay follows then from Little’s Theorem:

$$\bar{T}_{\text{delay}} = \bar{N} \cdot \frac{1}{\lambda}.$$ 

Note that the solutions of the GPS model, e.g. the sojourn time or job number distribution, are insensitive with respect to the job size distribution. More precisely, only the mean job size $1/\mu$ affects the results.

There is a special case of the GPS model: the ordinary PS model. If referring to this model, we assume that the users can utilize all available capacity, i.e. $r_n = C$. The distribution of $N$ is then simplified to:
\[ P(N = n) = \rho^n (1 - \rho) \text{ with } \rho = \frac{\lambda}{\mu C}. \]

For more details on Generalized Processor Sharing models, please refer to [Cohen].

**Ordinary PS**

From the packet level simulations we have observed that the “saturation throughput” of a WLAN does not alter significantly with different numbers of present upstream TCP flows, see section 3.2.2.2. If the data is transmitted in another direction (from the “web server” through AP to the WLAN users, downstream), the AP is the only STA in the BSS that has bulk data to send. Although there are many concurrent flows in the scenario, the WLAN will probably not behave much differently from if there is only a single flow present. The STAs only have occasionally TCP ACK packets to send and therefore only few stations are contending for the WLAN medium at any given time.

We have seen that TCP over WLAN allocates a fair share to the WLAN users concerning flows with finite file sizes. The GPS model also discusses the situation where server capacity is fairly shared among the jobs. Therefore we propose to utilize the results from the ordinary PS model, and take the effective medium throughput as the server capacity \( C \).

Because a WLAN service provider would not allow infinite number of users into the network. We also take the maximum number of users into consideration. [Cohen] provides extensive results on PS models and we will adapt the results for our specific system. Without further going into details, we provide the normalization operation below:

If there are at most \( m \) flows allowed simultaneously in an ordinary PS system, the distribution of the number of flows \( N \) becomes:

\[
P^n(N = n) = \sum_{t=0}^{m} P^n(N = t) = \begin{cases} 
1 - \rho \cdot (\rho)^n, & n = 0, 1, ..., m \\
1 - \rho^{m+1}, & 0 \\
0, & \text{elsewhere}
\end{cases}
\]

The mean number of flows in such a system is:

\[
\bar{N} = \sum_{n=0}^{m} [n \cdot P^n(N = n)] = \frac{1 - \rho}{1 - \rho^{m+1}} \cdot \sum_{n=0}^{m} [n \cdot \rho^n]
\]

If utilizing the result for finite harmonic series:

\[
\sum_{n=0}^{m} r^n = \frac{(1 - r^{m+1})}{(1 - r)},
\]

and the relation,

\[
n \cdot r^n = r \cdot \frac{dr^n}{dn}
\]

We obtain:

\[
\bar{N} = \frac{\rho(1 - (m + 1) \cdot \rho^n + m \cdot \rho^{m+1})}{(1 - \rho)(1 - \rho^{m+1})}
\]
The mean transfer time can then be calculated by applying Little’s Theorem:

$$\bar{T}_{\text{delay}} = \frac{1}{\lambda (1 - P^*(N = m))}$$

In order to utilize the GPS results, a number of simulations are conducted for more insight into the packet level characteristics of TCP over WLAN, including its saturation throughput against the number of active stations and the fairness on capacity share. The latter is essential to apply the GPS results.

### 3.2.2 TCP packet level studies

#### 3.2.2.1 Timing of TCP and WLAN interaction

Unlike UDP traffic, interaction between TCP source and destination plays an important role. This means that on both TCP and 802.11 MAC layers ACK packets are sent\(^{15}\). The WLAN MAC layer is not aware of the content of the data if it receives it from higher OSI layers. Hence TCP data packets and TCP ACK packets are treated equally. Because of the Backoff mechanism, it has been suggested that TCP data packets contend with its own ACK packets occasionally ([Wu]). This means that even with one single TCP connection through AP, there are collisions.

In order to gain more insight into the interaction between TCP and 802.11 MAC, and the way NS-2 interprets it, we have run a simulation with only one TCP-pair. The TCP connection is made between the AP and one wireless STA. Based on the trace file generated by NS-2, transmissions of the first a few TCP data packets and other relevant packets have been reconstructed. See Figure 3-6. In this simulation traditional TCP (Tahoe) and delayed acknowledgment are used, which can be found in a vast amount of applications on different platforms today.

![Figure 3-6 Interaction of TCP with WLAN MAC layer](image)

Packet transmissions are drawn above and receptions below the line at each layer.

It can be seen that post backoff takes place at (A) after the first data packet has been sent. At (B) the TX-queue in the sender is empty, at the receiver side a MAC ACK

\(^{15}\) To avoid confusion in the remaining of this document, we will denote ACK packets on WLAN MAC layer by MAC ACK and ACK packets in a TCP flow by TCP ACK.
segment is being transmitted while a TCP ACK is generated. After the successful transmission of the MAC ACK, the receiver waits more than a DIFS before it begins with the transmission of the TCP ACK packet. The interval is determined to be always a sum of the duration of a DIFS and a random number of timeslots. Apparently the medium is sensed busy because the receiver itself is sending a MAC ACK packet if the TCP ACK packet is offered to the MAC layer. At (C) the receiver has received again 2 TCP data packets and generated a second TCP ACK. In the mean time the sender also has a TCP-data packet queued in its TX-buffer. It can be seen from the trace file that the TCP data packet “wins” this contention and is granted to transmit first. Later at (D) the same situation occurs again if both the sender and the receiver have non-empty TX-queues. This time the receiver “wins” and hence the TCP ACK packet is transmitted first.

It has also been observed that the receiver’s MAC finishes delivering the content of a receiving MSDU after it sends the MAC ACK packet. The delay between the sending of the MAC ACK packet and receiving of TCP data packets on TCP layer is 15 μs. This length does not vary even if the payload size is changed; simulation with other TCP payload size has namely yielded the same result. For every other TCP data packet delivered to its destination, a TCP ACK packet is generated (delayed acknowledgement). This is done without any delay. The transmission of a MAC ACK takes 304 μs, as shown in the calculation in 3.1.1. Hence at the arrival of the TCP ACK packet on the TCP destination’s MAC layer, the medium is detected busy and therefore the TCP ACK packet is deferred. The same also takes place at the sender after reception of a TCP ACK packet, except that the TX-buffer at the sender in the most cases is already filled with data packets because of TCP’s open window.

3.2.2 TCP “saturation throughput”

The available capacity for TCP flows, i.e. the TCP saturation throughput, is an important parameter in the GPS models. In these simulations we consider a scenario as depicted in Figure 3-7. In our research files are transmitted between a STA and the “distant web server”. Compared to the UDP scenario in Figure 3-1, the additional items “web server” and “internet” can be found. With “Internet” we introduce propagation delay caused by an external network connected to the WLAN BSS.

TCP is a connection-oriented protocol. In its congestion control mechanism the feedback from the data destination is essential in determining the flow transfer rate. In the real world it occurs rarely that the wireless stations and the access point in a certain BSS form a TCP pair. The connection is usually established between a WLAN user (a STA) and a remote server, e.g. an FTP server, a mail server etc. Propagation delay in the external network has to be taken into consideration, for it affects one of the most important parameters on TCP performance, its Round Trip Time (RTT).
Relevant parameters in these simulations are shown in Table 3-2.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data rate</td>
<td>1 Mbps</td>
<td>TCP payload</td>
<td>1500 bytes</td>
</tr>
<tr>
<td>Basic rate</td>
<td>1 Mbps</td>
<td>IP header</td>
<td>20 bytes</td>
</tr>
<tr>
<td>SIFS</td>
<td>10 μs</td>
<td>TCP header</td>
<td>40 bytes</td>
</tr>
<tr>
<td>DIFS</td>
<td>50 μs</td>
<td>IFQ length</td>
<td>50 packets</td>
</tr>
<tr>
<td>EIFS</td>
<td>304 μs</td>
<td>RTS Threshold</td>
<td>300 bytes</td>
</tr>
<tr>
<td>MAC overhead</td>
<td>224 bits</td>
<td>Fixed link bandwidth</td>
<td>10/100 Mbps</td>
</tr>
<tr>
<td>PHY header</td>
<td>48 bits</td>
<td>Fixed link delay</td>
<td>10 ms</td>
</tr>
<tr>
<td>PLCP preamble</td>
<td>144 bits</td>
<td>TCP receiver $W_{\text{max}}$</td>
<td>20 packets</td>
</tr>
</tbody>
</table>

Table 3-2 Simulation parameters to determine TCP over WLAN saturation throughput

Although later in our research we will focus on the file transfer time in the downstream direction, we will first look at the saturation throughput with flows in the other direction, for this yields comparable results with the UDP scenario. If we consider the UDP scenario in download direction, it is also known that the saturation throughput does not decrease much with a larger number of stations, since the AP is the only active user competing for the WLAN medium. For TCP the same is expected: upload flows will probably behave more complex compared to download flows regarding the saturation throughput issue.

For each STA a TCP connection is set up with the distant web server via the AP. The simulation has duration of 2000 seconds. The data from the first 140 seconds are omitted to remove the transient effects and TCP slow start effects. The sources are started 200ms after each other, beginning with STA 1. The last procedure is to avoid the known TCP-synchronisation effect if numerous TCP flows have to share the same link ([Floyd]). TCP Tahoe and delayed ACK are used for these flows. The scenario parameters are set as shown in table. WLAN is operating in basic access mode. The simulation results of saturation throughput against the number of present stations/flows are shown in Figure 3-8.

---

16 This value is chosen such that the MPDUs containing TCP data will be transmitted through RTS/CTS mode, and the MPDUs containing TCP-ACK will be transmitted through Basic access mode.

17 In [802.11] section 7.2.2 it can be found that the MAC header omits field address 4 if transmission only takes place in the same BSS. This means that the overhead from the MAC layer is 28 bytes.

18 The fixed link should not be the bottleneck in the scenario; therefore we have taken 10 Mbps for WLAN at 1 Mbps and 100 Mbps for WLAN at 11 Mbps. The link delay is kept the same.
The throughput drops if more stations are added to the scenario like in the UDP scenario. The drop is very limited compared to the UDP scenario. The saturation throughput in the UDP scenario drops steadily if the number of stations becomes larger. In the TCP scenario this holds until five stations are present, after that the saturation throughput hardly drops. Because all TCP sources are sending towards the web server via the AP, all TCP ACK packets are also sent through the AP in the backward direction to the wireless STAs. The AP has to compete with all the stations in the scenario for medium access and therefore cannot deliver the TCP ACKs as fast as it receives from the WEB server to send. TCP ACKs may also drop if the buffer at the AP is full. Due to its congestion control mechanism, the TCP sources do not put packets persistently to send if they do not receive ACK for their previous packets. The consequence is that the number of sources, which have packets to transmit, remains low. The saturation throughput therefore remains at the level of a few contending stations. The small decrease at a larger number of stations can probably again be accounted by the fact that there are more routing packets sent.

The transmission rate for TCP over WLAN can also be calculated if there is one TCP connection active.

Since TCP Tahoe and delayed ACK are used, a TCP ACK packet is generated for every two TCP data packet. For each either TCP data packet or TCP ACK packet a MAC ACK packet is sent. So an ideal cycle contains two MAC data frames with TCP data, one MAC data frame with TCP ACK packet. Since there are two backoff counters at both the sender and the receiver, the backoffs can take place simultaneously, in the most optimistic scenario the ideal cycle contains two backoff windows and in the worst scenario three backoff windows. We take the average value of 2.5 backoff windows in our approximation:

\[
T_{cycle} = 2 \times T_{DATA} + T_{TCPACK} + 3 \times T_{MacACK} + 3 \times (SIFS + DIFS) + 2.5 \times T_{Backoff} + 6 \times T_{delay} \\
= 2 \times 12896 + 896 + 3 \times 304 + 3 \times (10 + 50) + 2.5 \times 310 + 6 \times 2 \\
= 28567\,\mu s
\]
There is also a probability that the forward TCP data packet collide with the backward TCP ACK packets. This probability is approximately 1/32 due to the fact that the random backoff window is uniformly distributed in [0, 31]. If collision takes place, the cycle is enlarged with the transmission time of an extra TCP data packet. The reason is that each collision always involves one data and one ACK packet. The time wasted is the time to have the largest packet completed, which is the data packet in this case. The backoff window is also increased to a double size in that case: U[0, 63] with an average of \( \frac{31 \cdot 20}{2} = 630 \mu s \). The length increased for collision is then

\[
T_{\text{collision}} = T_{\text{DATA}} + SIFS + DIFS + \overline{T_{\text{backoff}}} = 12896 + 10 + 50 + 630 = 13586 \mu s
\]

The probability of two successive collisions is considered negligible and is not taken into considerations in this approximation.

The rate at which TCP can send is at most

\[
R_{TCP,B} = \frac{2 \times 8 \times 15000}{28567 + 13586 / 32} = 828.032/1358628567 Mbps
\]

From simulations a saturation throughput of 0.809Mbps has been shown. Although the routing packets cause some overhead, the difference here is still too significant. In further research another factor has been found which contributes to the difference. We will give the possible main reason for the observed difference in section 3.2.4.

If data is transmitted using TCP over WLAN in RTS/CTS mode, more overhead is introduced due to the additional RTS and CTS packets. See Figure 3-10 (Note that the TCP ACK packet is not sent through RTS/CTS mode).

Simulation from NS-2 yields

We can calculate the theoretical maximum throughput for a single TCP flow. In RTS/CTS mode the ideal cycle contains a few more elements:
Collisions in this scenario involve a RTS packet and a MAC data frame containing a TCP ACK packet. In that case the ideal cycle is enlarged by:

\[ T_{\text{collide}} = \max[T_{TCP,ACK}, \text{RTS}] + SIFS + DIFS + \overline{T_{\text{Backoff}}} = 896 + 10 + 50 + 630 = 1586 \mu s \]

Hence, the maximum rate of one TCP flow in WLAN’s RTS/CTS mode is:

\[ R_{TCP,RTS} = \frac{2 \times 8 \times 1500}{29902 + 1586} = 0.801 \text{Mbps} \]

NS-2 simulations have shown a saturation throughput of 0.789 Mbps with one TCP flow.

### 3.2.2.3 Fairness

A condition to apply the PS model is to have the system capacity equally distributed among the users. Although TCP and WLAN are both designed for fair capacity share, little is known about the behaviour of TCP flows over WLAN. Therefore, the fairness issue is investigated.

In 3.2.2.2 we have found that the saturation throughput of TCP over WLAN does not drop as severely as UDP in the basic access mode. The possible explanation here is that even if the number of present flows is increased, the number of stations that are active is limited. Due to the ACK losses and its congestion control mechanism, TCP flows are not persistent as UDP controlled flows, in the sense that a TCP source might already have sent all the packets in its current window. It does not put the next packet into its TX queue before it receives another ACK from previously sent data packets.

If we examine the data traffic of each STA in a TCP upstream scenario, we see a typical distribution like shown in Figure 3-12: (the system has the same settings as described in section 3.2.2.2)

![Figure 3-12 TCP unfairness over WLAN](image)

The amount of traffic sent by each STA differs enormously. Some STAs can transmit at full speed while others can rarely send any packets. It shows clearly that not all STAs can send an equal amount of data. There is some randomness in which STAs can send at full speed and which not, although the ones that start first tend to be more “lucky” than the flows that started later.
In various researches it has been shown that TCP Tahoe allocates the available capacity to competing flows equally over wired lines (\cite{Stevens}, \cite{Hasegawa}). At first sight Figure 3-12 shows that WLAN seems to behave differently. In WLAN systems the medium is not full duplex, the forward TCP data packets have to contend with the backwards TCP ACK packets. The AP sends all TCP ACK packets over the WLAN medium while the STAs send the TCP data packets. Recall that in a legacy 802.11b system the AP has no priority with respect to sending data. Packetwise, the AP only gets equal share of the WLAN capacity as each of the rest of the containing stations. The TX queue at the AP towards the wireless stations grows fast if many upstream TCP flows are active. Because there is only a finite buffer available, backward ACK packets are dropped if many TCP flows have their window built up. The loss of ACK packets triggers the congestion control mechanism within TCP and decrease the transfer rate of the STAs.

It is known from TCP that if congestion takes place, flows with a larger open window are less sensitive to TCP ACK losses (\cite{Pilosof}). This is due to the cumulative acknowledgement nature of TCP whereby the next ACK packet has the appropriate sequence number and make up for the loss of the previous ACK packet (\cite{Stevens}). Flows with a larger open window have more ACK packets ongoing and therefore also a higher probability to acquire a position in the TX buffer at the AP. On the other hand, the loss of an ACK packet for a flow with a small window can be disastrous and lead to enormous timeouts due to its exponential growth.

In our scenario, the flows that are start earlier would have a relatively larger window built up than the latecomers if congestion takes place at the access point. This possibly gives the explanation why the first STAs tend to be “luckier” in acquiring the medium capacity. Due to the random nature of the backoff mechanism there is always a possibility that a few ACK packets are lost successively from a source that has built up a full window. This is possibly why the most advantageous STA 1 also could have a low good put in the example shown (Figure 3-12).

Hence this unfairness mainly comes from the finite buffer at the AP. To verify this result we have increased the buffer size exceeding the maximum number of possible ongoing ACK packets. Denote the buffer size at AP by $B$ and the TCP receivers advertised window size by $W_{\text{max}}$ and the number of simultaneous flows by $n$, we have chosen $B \geq n \cdot W_{\text{max}}$ to eliminate the loss chance.

If other parameters are kept the same, we obtained the results in Figure 3-13 with the enlarged buffer size.

![TCP "unfairness" with enlarged AP buffer](image-url)
This graph shows that the medium capacity is distributed considerably more equally among the TCP sources.

In our scenario with limited buffer size, if a new flow arrives at a (nearly) fully utilized WLAN medium, it is nearly impossible to compete with other flows running at full window size. Contrary to the infinite flows used in our scenario, real world TCP flows do not carry files of infinite size. This means that if a certain STA is at the advantageous positions and transmitting with $W_{\text{max}}$, it does not last forever. If the current file is finished and a next TCP connection at the same STA has to be set up, it will do its slow start again and put itself into the weak competing position. At the same time the departure of this flow also frees some link capacity such that other flows could gradually have improved transfer rate (larger window). The “fairness” shall improve if TCP carries finite flows. In Figure 3-14 TCP flows over WLAN have been examined that contain files of infinite sizes. These flows have deterministic sizes and are started directly after the previous one is finished.

The NS-2 simulation results show similar behaviour as we have predicted. The fairness “improves” if files of finite sizes are transmitted through TCP. The above figures show that the share is more equal with smaller files. This is due to the fact that within our simulation time of 2000 seconds, more files of 10 packets can be sent than files of 100 packets. Therefore the STAs have to set up a TCP connection and go through slow start much more frequently with files of 10 packets. Over a long period of time, TCP sources are expected to get equal share of the link capacity as long as the TCP connections are reset from time to time and their flow distributions have similar parameters.

3.2.2.4 Summary
TCP needs to receive Transport layer acknowledgment (TCP ACK) on the backward direction. If TCP file transfer over WLAN is considered, the forward TCP data packets contend the medium with the backward TCP ACK packets. NS-2 simulations confirm
this behaviour. NS-2 simulations also show that the WLAN medium is sensed busy if a packet from higher layer is offered to it while the station is still sending a MAC ACK packet.

Compared to the dependence of saturation throughput on the number of STAs in the UDP scenario, the values stay fairly stable for flows with TCP over WLAN. Although on flow level the number of TCP flows is high, i.e. there are TCP connections between the AP and each of the STAs; on packet level not all of these flows are simultaneously active. Therefore the real number of contending station is not as large as the number of present flows.

TCP flows, which have built up a larger window, are less sensitive to ACK losses. They likely acquired more available capacity than flows that did not build up their window. This is shown clearly if flows of infinite size are examined. In real world all flows will have limited size and restart in disadvantageous position (slow start) all the time. The medium capacity will be distributed equally among the sources.

3.2.3 Numerical flow level results of the integrated ordinary PS model

In the previous sections the important parameter for the ordinary PS model, the TCP saturation throughput, is determined. In addition, it is shown that the TCP flows fairly share the WLAN capacity on the long turn. This allows us to utilize the ordinary PS results to predict file transfer times with TCP over WLAN.

We have run some simulations with NS-2 2.27 to verify this simple model. The scenario used is shown in Figure 3-7. We have chosen to allow maximal \( m = 50 \) simultaneous flows into the system. The WLAN parameters are used as the packet level simulations, see Table 3-2, besides we have also run a few simulations with 11Mbps WLAN’s. In that case we choose for WLAN the parameters data rate 11Mbps and basic rate 2Mbps (802.11b multi rates support, see [802.11b] revised section 7.3.2.2 and 9.2)

The flows arrive according to a Poisson process and have an exponential distributed size with mean:

\[
\frac{1}{\mu} = \{10, 100, 1000\} \times 1500 \text{ bytes}
\]

The transfer times are determined for various system load, which is defined as:

\[
\rho = \frac{k}{dR_{\text{WLAN}}},
\]

with

\[ R_{\text{WLAN}} = \{1,1\} \text{ Mbps} \cdot \]

To avoid transient effects, the data is collected only if each STA has finished receiving at least 3 files. Afterwards we collect the file transfer times during 5000 seconds. Only completed flows are counted. Each simulation point consist 5-10 simulations in order to achieve a 95% confidence interval fraction ratio of below 5%. The results are shown below.
In Figure 3-15 we see the flow transfer times from both NS-2 simulations and the prediction from the ordinary PS model with different flow sizes. WLAN’s access speed is 1Mbps. We have determined in section 3.2.2.2 that the effective goodput for one single TCP flow is 0.828 Mbps and this value is used as $C$ to feed into the ordinary PS model. The model predictions do follow the shape of the simulation results. However, the ordinary PS model tends to give a too optimistic prediction. The difference is even larger if smaller files are transmitted. We will give an explanation later in this section.

Following the same heuristics shown in section 3.2.2.2 we can determine the effective goodput for one single TCP flow is approximately 5 Mbps for a WLAN with 11Mbps data rate and 2 Mbps basic rate. Figure 3-16 shows the results from both NS-2 simulations and the ordinary PS model with capacity 5 Mbps. The model shows also overall a too optimistic prediction. While the file size is large, the model gives an accurate prediction with simulation at low $\rho'$, although the model’s performance becomes worse if the system is approaching saturation (around $\rho' = 0.45$). If files of smaller sizes are transmitted, the model does not give an accurate approximation at all.

Generally speaking, the ordinary PS model is too optimistic for prediction of the file transfer times. It seems that the “constant” capacity, which we have assumed for the ordinary PS model, is overestimated. Apparently the simple ordinary processor-sharing model alone does not give a good estimation.

### 3.2.4 TCP file transfers over WLAN, packet-level revisited

As shown previously, taking a constant capacity as if there is one flow present is just too optimistic.

The saturation throughput of TCP (good put) in the downstream direction is determined with different file sizes. The flows have deterministic sizes, namely {$10, 100, ...$}
100)*1500 bytes. A flow is started immediately to a STA if the previous one is finished. Refer to Table 3-2 for other simulation parameters.

Figure 3-17 Downstream TCP saturation throughput over WLAN at {1Mbps, 11Mbps}

The results show a number of interesting things:

1. If TCP flows contain only small files, they cannot utilize the full medium capacity. Due to slow start effect the files are already finished before the flow can reach its maximum window. This effect is even more visible if WLAN at higher rate is taken into consideration.

Denote the round trip time of the TCP flows by $RTT$, and the file size by $S$ (packets of 1500 bytes). Now consider a TCP flow with delayed acknowledgement and the flow initiates with 2 successive DATA packets. If there is no limitation on window size, the window will double its size after each RTT, i.e. 2, 4, 8, …. It will take approximately $\lceil \log_2(S + 1) \rceil$ number of RTTs to finish this file. Therefore, the rate that this flow can achieve is:

$$R_s = \frac{S \cdot 1500 \cdot 8}{\lceil \log_2(S + 1) \rceil \cdot RTT},$$

$R_s$ is an increasing function in $S$. In the simulation results in Figure 3-17 we see indeed that the small file size $S$ could impose a limited $R_s$, which is considerably lower than the medium effective capacity allowed by the WLAN medium. This explains also partially why the ordinary PS model gives a very inaccurate estimation for small files.

2. Files with 1000 packets do not suffer much from the effect described in 1. However the difference between the maximum aggregate throughput with 2 flows and 50 flows is still 10%. Because NS-2 only shows TCP good out, the loss of capacity might come from the limited downstream buffer at the access point. If there are many concurrent flows, the total window size will exceed the given buffer size and the
result is that part of the TCP data packets is lost. If we look at the TCP good put with an enlarged buffer, which is sufficient to contain ongoing packets if all 50 flows are at their max window, we see that the aggregate throughput is much more stable. See Figure 3-18 (with a sufficient buffer no loss will occur and thus the total bytes transmitted will be equal to good put). If there is only a buffer for 50 packets, the good put decreases, especially if a large number of concurrent TCP flows are active. It is interesting to see that in that case the TCP sources actually send more packets. However, a part of these packets are dropped due to buffer overflow. The loss increases with the number of concurrent flows.

![Figure 3-18 The influence of AP buffer on TCP good put](image)

The good put achieved by one TCP flow is 0.830 Mbps, which is very close to the outcome of our calculation in section 3.2.2.2. This shows that the assumption of a constant capacity may still hold, although we have to take the loss at the AP due to limited buffer size into consideration.

### 3.2.5 Refined TCP packet/flow level integrated model

From previous results it seems that WLAN provides an equal capacity for various numbers of simultaneous TCP flows. This suggests that we could use TCP results for fixed links to model the transfer time with TCP.

Lassila et al. ([Lassila]) have proposed an integrated packet/flow level model for file transfer time predictions with TCP over fixed link. At flow level the authors have used the same results from a Generalized Processor Sharing model as shown above. At packet level the model is as follows:

It is assumed that there are a fix number of flows actively transmitting over a shared link. Furthermore these clients will be assumed to have identical RTTs. It has been shown by many studies that TCP’s traffic load and the loss probability are related by a squared root formula. The load equation from Kelly ([Kelly]) is used to represent this relation:

\[ t_n = \frac{n}{RTT} \sqrt{\frac{2(1-p)}{p}} \]

Where \( n \) is the number of flows, \( p \) the loss probability and \( t_n \) the offered load if \( n \) flows are active. While the equation above shows the dependence of \( t_n \) on \( p \). The loss probability \( p \) is also dependent on the traffic load which might be described by a queueing model. In Lassila’s approach they have chosen to use an M/D/1/K queueing model (M: Markovian, Poisson job arrivals; D: deterministic service times; 1: single server; K: limited queueing buffer K). The packets arrive following a Poisson arrival
process and the service time is constant and equal to 1. The system load is $t_n$. There are a few methods to compute the loss probability $p$ from the system described above (see e.g. [Virtamo], [Tijms]). Lassila et al. have assumed from the M/D/1/K results that $p$ is a monotone increasing function in $t_n$ while the formula from Kelly shows that $t_n$ is monotonously decreasing in $p$. These relations guarantee the existence of a unique fixed point solution.

According to the heuristic described in [Lassila] the system properties RTT and buffer size K are used as input. Based on these parameters the fixed point equations shall be solved to obtain valid values for $p$ and $t_n$. The aggregate throughput at packet level, given that there are $n$ flows present, is described by the relation $s_n = t_n \cdot (1 - p)$. Subsequently, the GPS equations at flow level can be solved by substituting $r_n$ (joint transmission rate of $n$ simultaneous flows) with $s_n$. If the mean number of flows is known, we can use Little’s Theorem to calculate the mean file transfer time.

In [Lassila] the authors have considered TCP flows from access links with limited capacity. It is not be relevant to our work and hence not described here. If a large buffer is present, the queuing delay can increase if more flows are competing for the link. This can become a significant part in a TCP flow’s RTT in contrary to the assumption made in [Kelly]. In [Lassila] the RTT definition is hence adjusted to $RTT_n = RTT_0 + d_n$ with $RTT_0$ representing transmission delays from the links and $d_n$ the mean queuing delay caused in the M/D/1/K queue. This adjusted $RTT_n$ is used in the fixed point equations described above. Another refinement made by Lassila et al is the TCP slow start compensation. The above equation assumes that TCP flows can transmit at full window size in the entire transmission period and yields usually a too optimistic estimation of the file transfer delay. In realistic scenarios TCP flows will have much fewer packets transmitted during slow start. To make compensation of this effect Lassila et al have calculated the number of packets $b$, which is the difference between a TCP flow sending with and without slow start taken into consideration. The necessary time to transmit these packets is added to the results before. Expressed in formula:

$$ \bar{T}_{\text{delay}} = \frac{N}{\lambda} + b \cdot \frac{RTT(\hat{t})}{w} $$

In which $\bar{T}_{\text{delay}}$ is the total mean transmission delay, $\bar{N}$ the mean number of flows in the system, $\lambda$ the flow arrival rate, $b$ the “unsent” packets as consequence of TCP slow start, $RTT(\hat{t})$ the refined round trip time of a TCP flow, $w$ denotes the mean window size of the TCP flows in the equilibrium state.

In our scenarios we can calculate $RTT_0$:

For the 10Mbps/1Mbps wired-cum-wireless network (10ms delay in fixed link):

$$ RTT_0(1Mbps) = T_{\text{wired}}(\text{data}) + T_{\text{wireless}}(\text{data}) + T_{\text{wireless}}(\text{ACK}) + T_{\text{wired}}(\text{ACK}) \approx 0.036\text{s} $$

and for the 100Mbps/11Mbps wired-cum-wireless network (10ms delay in fixed link):

$$ RTT_0(11Mbps) = T_{\text{wired}}(\text{data}) + T_{\text{wireless}}(\text{data}) + T_{\text{wireless}}(\text{ACK}) + T_{\text{wired}}(\text{ACK}) \approx 0.023\text{s} $$

In the next figures the results from our simulations and the predictions from Lassila’s model are represented.
We see that the model from Lassila does slightly better than the ordinary PS model we have used before. Due to the slow start correction the model from Lassila gives a better prediction at lower \( \rho \)'s. At higher \( \rho \)'s the results from Lassila is similar to the ordinary PS model. In that region there are many concurrent flows in the system and the average window size is small, therefore the slow start compensation only has a minor impact on the total transfer time. However, the mean transfer time is still higher than the ordinary PS model, hence closer to the simulations results. This can possibly be explained that the integrated model takes into account the loss in TCP-flows due to the limited amount of AP buffer and therefore yields a lower and hence better prediction for the effective WLAN throughput.

Overall the Lassila model still gives a too optimistic prediction. The reason here might be the optimistic assumption that the TCP packets arrive according to a Poisson arrival process. In reality the packet arrival process might be much bustier than can be described by an M/D/1/K queue. Related research for improvement of the packet level approach can be found in [Haan].

To verify our assumption that TCP file transfer times over WLAN can be approached by TCP over a wired link with the WLAN effective capacity, we have also conducted simulations with the WLAN substituted by a fixed link with WLAN’s effective throughput at both 1Mbps and 11Mbps. Other parameters are kept the same. In Figure 3-19 we see that the outcome from the fixed link simulations almost exactly fit the curves from our WLAN simulations. This suggests that TCP models for fixed links shall also be valid for a WLAN, as long as the effective throughput of TCP over WLAN can be determined.

### 3.2.6 Summary

In this section we have applied the ordinary Processor Sharing model and an integrated packet/flow level model to model the file transfer time using TCP over WLAN. The results of the models show similar shape as the simulations. They also show the same critical region where the transfer times grow rapidly. However, the models provide a too optimistic prediction of the transfer times. The main reason is that the approximation of the medium throughput is too optimistic. We have also seen that the behaviour of file transfers with TCP over WLAN, in the downstream direction, is very similar with a fixed LAN with a capacity equal to the effective throughput of WLAN with one TCP flow. The latter can be approximated with a simple calculation. Improvement of the file transfer time model shall be obtained through further refinements in the TCP models. The packet arrival process shall be improved and the TCP throughput of small files should also be taken into consideration as addition to the integrated models.
3.3 Conclusions

In this chapter we have studied the behaviour of file transfers with TCP over WLAN and used two models to predict file transfer time with TCP over WLAN.

To obtain reliable simulation results, we have first examined the simulation package Network Simulator 2 (NS-2). We determined the “saturation throughput” with persistent UDP flows both through NS-2 simulations and mathematical models from Wu ([Wu]). From the results we see that NS-2 delivers fairly accurate results. At high number of concurrent flows the model and NS-2 do deliver different results. The main reason here is due to the fact that [Wu] has not considered the influence from EIFS while EIFS is implemented in NS-2.

Secondly we investigated the behaviour of TCP over WLAN at packet level. Simulation with TCP flows of unlimited sizes delivers inaccurate results toward fairness since the TCP flows with a larger window have a larger advantage in obtaining the medium capacity. If flows are of infinite size, flows have very small probability to increase its window and get a fair share from the medium capacity. If flows of finite size are used, TCP does distribute the medium capacity fairly among the flows.

In the third place the flow level performance of TCP file transfers over WLAN are examined. We have studied flows with exponential distributed interarrival times (Poisson arrival process). Two models have been examined: the ordinary Processor Sharing model with a fixed capacity and an integrated packet/flow level model proposed by Lassila et al ([Lassila]). Although the latter gives a better prediction of the flow transfer time, they both give too optimistic predictions. At flow level TCP file transfers over WLAN show similar behaviours as TCP file transfers over a fixed link. Thereby the capacity of the fixed link is taken as the effective throughput of the compared WLAN.

Hence, in order to improve file transfer time predictions with TCP over WLAN, better TCP models have to be used. Based on literature ([Lassila], [Haan]), it is usually too optimistic to use M/D/1/K queue to model TCP packet losses at packet level. Hence, the resulting medium throughput is also too optimistic. Generally the arrival process of TCP packets is believed to be much bustier and cannot be approached by a Poisson arrival process like described in an M/D/1/K queue. Other refinements might include the consideration of the achievable rate of TCP flows with very limited size. Due to slow start effect these flows cannot reach their maximum allowed rate by the link. Our simulations also confirmed this. If this effect is taken into consideration, the rate to apply the PS solutions at flow level is expected to be much lower at low system load. The results should yield a higher transfer rate and give a more accurate prediction. Another approach is to utilize state dependent PS models. Thereby the “saturation throughput” of TCP shall be determined for different numbers of concurrent flows.
Wireless Local Network 802.11e Simulation Studies

This section examines the upcoming new IEEE standard: 802.11e. As explained in the second section, IEEE 802.11 Task Group e is set up to integrate QoS enhancement into 802.11. Simulations are used to determine the performance of this new member of the 802.11 family.

This section is organized as follows: section 4.1 describes the implementation of the 802.11e extension to NS-2. Section 4.2 presents the results of performance evaluation of the different parameters within 802.11e. In section 4.3 the performance of 802.11e is examined in a scenario where various services are present, including VoIP, VoD and Web browsing.

4.1 Extensions of NS-2

802.11e is a new technique and not included in the official releases of NS-2. However, over the last few years various research groups have developed their own 802.11e extensions for NS-2, e.g. Atheros ([NSe-a]), University of Stanford ([NSe-s]), INRIA ([NSe-i]). Another popularly used package is from the Telecommunication (TKN) group of the Technical University of Berlin ([NSe-b]). The latter is based on the latest 802.11e draft and built for the latest NS-2. Because it contains the most up to date information from the draft to the author’s knowledge, the simulations from this chapter are based on the TKN implementation. Note that TKN has implemented AIFS as:

\[ AIFS_{TC} = SIFS + AIFS_{TC} \times \text{TimeSlot}, \text{with } AIFS_{TC} \in \{1, 2, 3, \ldots\}, \]

and TXOPLimit is defined in time.

4.1.1 Code walk through

Within the TKN codes all the four parameters of 802.11e are implemented, as well as the persistent factor that is not longer included in the 802.11e drafts. The persistent factor is not considered in the remaining of this section. The parameter values are stored in the arrays \(cwmin\), \(cwmax\), \(aifs\) and \(txop_limit\). For example, \(cwmin_f0\), \(cwmax_f0\), \(aifs_f0\) and \(txop_limit_f0\) correspond to the parameter set of traffic class TC[0].

Each TC has its own queue within a station with the pre-defined parameter values. Up to four customizable queues are defined in the TKN package. Although this is less than specified in IEEE 802.11e draft, in reality few networks would require more than four TCs.

The original NS-2 2.26 includes a set of timers for the queue in a station, i.e. defer timer, backoff timer, interface timer, send timer and NAV timer. These timers are responsible to handle the backoff process of the 802.11 stations. Within the 802.11e package these timers are linked to the TC queues instead of the stations. The backoff process is now done on the queue level as specified in the 802.11e draft. The backoff processes of these queues are determined by the TC parameters \(cwmin\), \(cwmax\) and \(aifs\). The contention mechanism is altered such that a packet of the highest TC within a station will be transmitted if collision takes place.

The duration of a transmission opportunity is defined with \(txop_limit\) of TC[0]. By default TXOPLimit is turned off, e.g. all queues are only allowed to transmit one
single packet after winning a contention. If TXOPLimit is turned on, the `txop_limit_[pri]` values apply. A maximum number of packets of the same queue are sent of which the total duration does not exceed the `txop_limit_[pri]` value. Besides the transmission time of the packets, different IFS intervals and MAC ACK transmission times are also taken into account.

4.1.2 Running simplified cases

As a basic test for the TKN package we have determined the UDP upstream saturation throughput for different numbers of stations. Thereby the 802.11e parameters have been assigned values that are the same as an 802.11b WLAN, namely \(\{\text{AIFS} = 2, \text{CW}_{\text{min}} = 31, \text{CW}_{\text{max}} = 1023, \text{TXOPLimit} = 0\}\). Other simulation parameters are set similar as used in (the previous chapter). The results are shown in Figure 4-1.

![Aggregate throughput vs. number of UDP sources](Upstream, 1Mbps WLAN)

Figure 4-1 Upstream saturation throughput with the TKN extension

In Figure 4-1 the saturation throughput, which are determined with NS 2.27 in the previous section is also plotted, together with the analytical results of Bianchi’s model. The results of simulations based on TKN implementation match almost exactly the reference NS 2.27 results.

By examining the codes it shows that TKN implementation includes all important features of IEEE 802.11e WLAN. The results of the simple test also indicate that it has a good working basic access mechanism.

4.2 Investigation of the impact of the parameters in IEEE 802.11e

IEEE 802.11 TGe is formed to integrate QoS enhancement on the WLAN MAC layer. Section 2 describes four variable parameters that have been introduced, namely AIFS, CWmin, CWmax and TXOPLimit (burst size). In order to assign proper values to these parameters in a specific 802.11e network, the impact of these parameters is first separately examined.

4.2.1 Discussion of the IEEE 802.11e parameters

- CWmin

[Bianchi] showed that the collision probability is greatly determined by the CW size at the contending stations. In 802.11e, different TCs have different CWmin values. Accordingly, the initial CW states among stations from different TCs are also different. Stations with a smaller CW value will have a higher probability to transmit. Therefore,
if the other parameters are set equal, stations with a smaller CW value will have a larger share of the medium capacity.

If a certain station has to draw a backoff window, the size of the CW is initially determined by its CWmin value. Hence the impact of CWmin will always be present. A smaller CW value means that the number of time slots in the backoff process is also smaller. The result is that less time is spent on the backoff process and therefore a positive impact is expected on the channel efficiency. However, smaller CW values also result in an increased probability of collisions. The second effect will become more dominant in networks with a high number of contending stations.

- CWmax

CWmax also contributes to the evolution of the CW in the contending process. Stations with smaller CWmax values are expected to obtain a larger share of the medium capacity. However, as opposed to CWmin, the value of CWmax is only reached after a number of successive collisions with the same packet. This suggests that differentiated CWmax values will only have effect with a relatively high number of contending stations. When a number is “relatively high” depends on the exact values of CWmin and CWmax, together with other parameters.

The two effects on channel efficiency described at CWmin are also valid for CWmax. However, CWmax defines the final CW value of its exponential growing process. Therefore, the value of CWmax is even more critical if collisions take place frequently, such as in a network with a high number of contending stations. A small CWmax value can greatly downgrade the system performance on channel efficiency if the number of contending stations is high.

- TXOPLimit

The TXOPLimit defines the duration that a STA is allowed to transmit after it has won a contention. Assuming that STAs from different TCs have the same parameter values except having different TXOPLimit value, the probability of a successful transmission will be the same for all these STAs. However, a STA from the traffic class with a higher TXOPLimit value will be able to transmit more data packets in one TXOP. From a throughput point of view, stations with a higher TXOPLimit value will obtain a larger share of the medium capacity.

With respect to TXOPLimit, the ratio of the throughput of a node in a network consisting of two TCs can be calculated based on [Bianchi]:

Define $N_1$ as the number of TC0 stations and $N_2$ the number of TC1 stations in the network, then the aggregate throughput of TC0 and TC1 is

$$RA_i = \frac{N_1}{N_1 + N_2} \cdot p_s \cdot \overline{B_i},$$

and
The ratio of the throughput of one single station of each TC is then:

\[
\text{ratio} = \frac{RA_1}{RA_2} = \frac{B_1}{B_2},
\]

which is the ratio of the payload sizes that the stations can transmit in their TXOPLimits.

- **AIFS(N)**

The length of AIFS determines after how many idle slots a station is allowed to count down its (residual) backoff count. If the other parameters are same, stations with a shorter AIFS will be able to start decreasing their backoff count earlier when the medium is sensed idle. This means that stations with smaller AIFS value will have an advantage in the backoff process. Hence, stations with lower AIFS are expected to get a greater share of the medium capacity.

Furthermore, the length of AIFS for a station does not alter in each backoff session. In a network where many stations are contending for medium access, it will take a station a few backoff sessions before it can transmit a packet. For a station with a larger value of AIFS, in each of these sessions it has to wait for a longer AIFS again before it can decrease its backoff count. The disadvantage for this station increases with the number of backoff sessions for sending each packet. And if a station with a large AIFS value has to contend with a large number of stations with small AIFS value, it is possible that this station can hardly decrement its (residual) backoff count. Upon expiration of its AIFS one of the other stations could already be transmitting. Hence differentiation through AIFS is expected to have a larger impact in a more crowded network.

### 4.2.2 Scenario description

The impact of these parameters is studied by means of throughput of STAs of different TCs in a saturated 802.11e network. The same scenario is applied as depicted in Figure 3-1. However, the stations are now divided into two TCs with different 802.11e parameter sets. We have chosen to evaluate a greater range of parameter values than the preferred ones in the 802.11e drafts, since the manufacturers and the operators are free to choose their own parameter values in their products.

In each scenario two TCs are present with different sets of parameter values. TC0 corresponds to a high priority class and TC1 corresponds to a lower priority class. Two types of scenarios are studied:

1. Same numbers of STAs for both TC0 and TC1, the total number is increased.
One TC has a fixed number of two STAs and the number of STAs from the other TC is increased.

In both scenarios the throughput per STA from different TCs is determined and compared.

In all scenarios the traffic is generated in the upstream direction and contains UDP packets with payloads of 1460 bytes. The contending stations are assumed to be persistent all the time, i.e. they always have packets to send. Our WLAN operates in its basic access mode with an access rate $r_{\text{WLAN}} = 1 \text{ Mbps}$.

A number of simulations are conducted to compare the impact of the 802.11e parameters, in these cases the TCs are differentiated with two different 802.11e parameters. In all simulations only the 802.11e parameter under investigation is varied, the other variable parameters are set according to their 802.11b equivalents, unless mentioned otherwise. The impact on differentiation is discussed per parameter below, as well as their effect on channel efficiency.

### 4.2.3 Simulation results

The simulations are performed with NS 2.26 with TKN’s 802.11e extension. The data from the first 140 seconds simulation time is omitted to avoid transient effects. After that statistics are collected for 2000 seconds simulation time. Simulation results are based on 4-6 replications in each scenario. The 95% confidence interval fraction ratio is below 5%.

1 CWmin

Simulations from type I scenarios yield the following results, see Figure 4-2. It shows the throughput of a single station in each TC, as well as the aggregate throughput of each TC. The ratio of the throughput per station of the two TCs is also plotted.

![Figure 4-2 IEEE 802.11e. Differentiation through CWmin. Type I scenario.](image)

(a) Left, TC0 with CWmin 7 and TC1 31; (b) right, TC0 with CWmin 7 and TC1 63

The results clearly show that stations with a smaller CWmin value obtain a larger share of the channel capacity than the other stations. If the total number of stations increases, the throughput per node decreases rapidly. The total amount of effective channel capacity drop due to increased number of collisions, and the decreased channel capacity is shared among a larger number of stations. An interesting observation is that the aggregate throughput of the lower TC stations is fairly insensitive to the number of stations in this type of scenario. Further investigations are required to provide more insight in this matter.
We see that the throughput ratio between stations from TC0 and TC1 (the green dashed line) decreases if more stations are contending for the medium. This suggests that differentiation through CWmin is less effective in a more crowded network, which confirms our expectation. It can also be observed that the throughput ratio lowers to the ratio of the CWmin values from different TCs, and will possibly even drop slightly below it if the number of stations further increases. This means differentiation through CWmin will always have impact.

Simulations from type II scenarios show similar behavior of a network containing stations with different CWmin values. The throughput per station ratio also drops to around the ratio of CWmin in these scenarios.

2 CWmax

Results from type I scenario simulations are shown in Figure 4-4.

In cases of differentiation through CWmax, the aggregate throughput of the higher priority TC is more constant. In Figure 4-4(a) the aggregate throughput of TC0 even increases until the total number of stations reaches 16.

The throughput ratio between TC0 and TC1 increases with the number of contending stations in the network. In a network where few stations are contending for the medium, the throughput ratio approaches 1, which means that there is practically no differentiation between the two TCs. This confirms our expectations.

In the left picture TC0 has a CWmax value of 31, which is equal to the CWmin value and hence no CW growth takes place. It’s interesting to see that there is
already an observable difference in the throughput of the two TCs when each TC only has one station in the network. Apparently the collision probability cannot be neglected even with two contending stations.

The impact of CWmax is stronger if the network has a larger number of stations. The consequence is that the traffic class with a lower priority, which means a larger CWmax value, can hardly get any medium capacity in a crowded network. The latter is often referred as *starvation* of the lower TC.

![Figure 4-5 IEEE 802.11e. Differentiation through CWmax. Type II scenario. TC0: CWmax 31; TC1: CWmax 1023](image)

(a) left, fixed number of TC1 stations; (b) right, fixed number of TC0 stations

Type II scenario shows similar results. The initial increase of the aggregate TC0 station throughput in Figure 4-5 (a) is because the increased number of TC0 stations, while TC1 has a fixed number of two stations. Due to the increased number of collisions the throughput drops after the total number of stations exceeds 8.

The throughput ratio starts again from one in a scenario where few stations are contending for the medium. The throughput ratio increases rapidly with the amount of contending stations. If the total number of stations is same, the ratio is higher if there are more stations of TC0 present.

The effect that the lower TC stations can hardly acquire any medium capacity can be observed again very clearly in Figure 4-5 (a).

3 TXOPLimit

Figure 4-6 and Figure 4-7 shows the throughput comparisons of differentiation through TXOPLimit in type I scenarios:

![Figure 4-6 IEEE 802.11e. Differentiation through TXOPLimit. Type I scenario](image)

(a) left, TC0 with TXOPLimit 0.1s and TC1 0;
(b) right, TC0 with TXOPLimit 0.1s and TC1 0.03s
In the type I scenarios we see a constant ratio between the throughput from the TC0 stations and the TC1 stations. In our simulations the data payload of each packet is 1460 bytes. When these packets are transmitted through a WLAN with an access rate of 1 Mbps, stations with a TXOPLimit of 0.03s are able to transmit 2 packets after winning a contention and a TXOPLimit of 0.1s allows a station to transmit 7 packets consecutively. In Figure 4-6 the ratio of the throughput of stations from TC0 and TC1 is almost constantly at 7 in (a) and 3.5 in (b), which are also the ratio between the packet counts of the TXOPLimit from the different TCs. Again, the expectations are confirmed here. Note that the ratio of throughput between stations from TC0 and TC1 increases slightly if a larger number of stations is contending for the WLAN medium. This differs from the discussion of TXOPLimit in section 4.2.1. Although code investigation does not show any strangeness, further investigation in TXOPLimit and the implementation is required to provide an explanation.

In type II scenarios similar results are obtained. The ratio in throughput follows the packet count in each TXOP from TC0 and TC1.

4 AIFS
Simulations of type I scenarios have yield the following results. See Figure 4-8.

Similar to differentiation through CWmax, the aggregate throughput of TC0 is fairly constant in Figure 4-8 (b) and even increases with the number of stations in Figure 4-8 (a).

Differentiation with AIFS always has impact on the throughput of stations of different TCs. The difference between the achieved throughputs of stations is little
with few contending stations and it increases steadily with the number of contending stations in the network. We also see that a small differentiation in AIFS already results in a great differentiation in the throughput performance, especially in a network with a large number of contending stations. Similar to the CWmax scenarios, the stations of TC1 also suffer from the starvation effect.

In Figure 4-9 the results of type II are displayed. The same observations can be made compared to type I scenarios. The throughput ratio increases with the number of contending station in the network. Figure 4-9 (a) shows the stations of high priority TC can retain a fairly stable throughput, regardless of the number of stations from the lower priority traffic class.

The throughput ratio lines in both Figure 4-9 (a) and (b) have a linear shape and match each other almost exactly. However, this is not the result of the chosen AIFS values, because in Figure 4-8 (a) the same AIFS combination is chosen while the ratio there has overall a larger value. The linear relation still holds. Further investigations are required to gain more insight.

5 Differentiation through two parameters

In order to compare the effectiveness of the IEEE 802.11e parameters, a number of simulations are conducted with differentiation in two parameters. Since the parameter TXOPLimit does not affect the transmission and collision probabilities, it is not considered here.

The TCs in these simulations are named after the parameter that is altered compared to 802.11b, e.g. in Figure 4-10 (a) the medium is shared by STAs of TC CWmin and TC CWmax. Note that in these simulations the examined parameters are always assigned a lower value compared to their 802.11b equivalents.
Figure 4-10 IEEE 802.11e. Differentiation through two parameters. Type I scenario. (a) above, CWmin and CWmax; (b) middle, AIFS and CWmin; (c) below, AIFS and CWmax.

Figure 4-10 (a) shows the results if both CWmin and CWmax are applied in the same 802.11e network. If there are a small number of STAs active in the network, a lower CWmin value (15 for one TC and 31 for the other) works more effectively compared to a lower CWmax value (127 to 1023). The mutual relation shifts more to CWmax’s favour if the number of active STAs increases. With a sufficient number of active STAs, STAs of TC CWmax obtain a larger throughput than STAs of TC CWmin. This confirms our findings that although differentiation with CWmin is always present, its effectiveness is less than CWmax in a crowded network.

In Figure 4-10 (b) the effectiveness of AIFS and CWmin is examined. While STAs of TC CWmin clearly obtain a larger share of the channel capacity with only a few active STAs, the share of the STAs of TC AIFS rapidly increases in a more crowded network. The aggregate throughput of STAs of TC AIFS even increases with the total number of active STAs. This shows that STAs with a lower AIFS are more effectively protected in a system with a large number of contending STAs.

Both CWmax and AIFS provide more effective protection with a larger number of contending STAs. Figure 4-10 (c) shows the results if these two parameters are applied for different STAs in the same network. Overall, STAs of TC AIFS have a
larger share of the channel capacity and their share relatively increases if the total number of STAs increases. However, the throughput ratio does not vary as much as in the other simulations. With carefully assigned AIFS and CWmax values, a network should always be able to provide more bandwidth to high priority users while preventing low priority users from starvation.

6 Medium efficiency

The total aggregate throughput of the previous scenarios are obtained by accumulating the aggregate throughput of the TCs and compared to a network containing only 802.11b stations. The latter is studied in section 3.1.1. Only the Type I scenarios are considered. The results are displayed in Figure 4-11:

![Figure 4-11 Impact of IEEE 802.11e parameters on medium efficiency.
(a) above left, CWmin; (b) above right, CWmax; (c) down left, TXOPLimit; (d) down right, AIFS](image)

In Figure 4-11 (a) we find that reduced CWmin value has a positive effect on the medium efficiency with few contending stations. If the number of contending stations increases, the aggregate throughput decreases and drops below the values of a network containing only 802.11b stations. This is inherent from the property of CWmin, see section 4.2.1. The increased probability of collisions obviously has more impact in case of a high number of contending stations. An increased CWmin value will have the opposite effect. We see indeed that the network containing both reduced and increased CWmin value stations performs slightly better than the one containing reduced CWmin value stations and 802.11b stations.

Similar effects can be observed for CWmax. However, the impact of CWmax can only be observed if there are a significant number of collisions such that CWmax can be reached. This is the reason that there is hardly any increase in the achieved aggregate throughput in the network with few contending stations. The negative impact of reduced CWmax value is indeed larger in a crowded network than a reduced CWmin value. A network containing stations with a larger CWmax value has a better performance in medium efficiency. This is shown in Figure 4-11 (b).
If TXOPLimit is applied, we observe an overall improvement in the medium efficiency, regardless of the number of stations in the network. Larger TXOPLimit value provides a higher medium efficiency. While the collision probability does not alter in a contention, more packets/data could be transmitted after a STA has successfully contended for the WLAN medium. For sending the same amount of data, a smaller number of contentions are conducted. The more packets a station is allowed to transmit in one TXOP, the better the throughput performance is of the network. This is an inherent property from TXOPLimit.

The last 802.11e parameter is AIFS. A shorter AIFS reduces the average time between the contentions and this has a positive effect on the medium efficiency. A longer AIFS is expected to affect the throughput performance in the opposite way. However, our simulations shows that a network containing two traffic classes, one with 802.11b stations and the other one with altered AIFS values, always performs better in medium efficiency compared to a homogeneous 802.11b network. When two traffic classes with different AIFS are present in the network, the stations from TC0 (higher priority) initially only contend with each other as their AIFS will expire earlier than the other class. This means that the concurrent number of contending stations is smaller than the total number of present stations. This effect is stronger when the number of stations is large in the network and the difference in AIFS value is larger between the different traffic classes.

### 4.2.4 Summary

In this section the impact of the 802.11e TC parameters has been examined by means of throughput analysis. It is shown that the differentiated parameter values result in different throughput at the different stations.

The impact on throughput differentiation is summarized in Table 4-1:
**Throughput differentiation grade**\(^{19}\)

### Network with a small number of contending stations

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CWmin</td>
<td>High. The effect is clearly observed and smaller CWmin value provides the corresponding TC with a high throughput compared to the lower TC. The ratio of throughput of different TCs is much higher than the CWmin value ratio.</td>
</tr>
<tr>
<td>CWmax</td>
<td>Very low. The effect of different CWmax values is hardly present in this region.</td>
</tr>
<tr>
<td>AIFS</td>
<td>Low. The impact of differentiated AIFS values can be observed. However, the differentiation is as strong as with CWmin and stations of lower TC can still get a fair share of the channel capacity.</td>
</tr>
<tr>
<td>TXOPLimit</td>
<td>Medium. The throughput ratio is much higher than the CWmin value ratio.</td>
</tr>
</tbody>
</table>

### Network with a large number of contending stations

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CWmin</td>
<td>Medium. Compared to its impact in a network with few contending stations, the differentiation grade decreases. However, the throughput ratio will not drop below the ratio of the CWmin values. Differentiation through CWmin will always have impact.</td>
</tr>
<tr>
<td>CWmax</td>
<td>High. The differentiation grade increases rapidly with the number of contending stations in the network. Stations of the lower priority TC suffer from “starvation”.</td>
</tr>
<tr>
<td>AIFS</td>
<td>High. The differentiation grade increases rapidly with the number of contending stations in the network. Stations of the lower priority TC suffer from “starvation”.</td>
</tr>
<tr>
<td>TXOPLimit</td>
<td>Medium. The throughput ratio follows almost exact the ratio of the TXOPLimit length.</td>
</tr>
</tbody>
</table>

Furthermore the impact of these parameters on the channel efficiency is investigated. This is done by assuming that two TCs are present in the network, with one TC consisting of 802.11b stations while the other one has a higher priority by altering one of the four parameters.

The impact of these parameters is shown in Table 4-2.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Effect on channel efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Small number of contending STAs</td>
</tr>
<tr>
<td>CWmin</td>
<td>Slightly positive</td>
</tr>
<tr>
<td>CWmax</td>
<td>Hardly present</td>
</tr>
<tr>
<td>AIFS</td>
<td>Hardly present</td>
</tr>
<tr>
<td>TXOPLimit</td>
<td>Positive</td>
</tr>
</tbody>
</table>

\(^{19}\) The grade is defined high if the differentiation works better for the higher priority class, which means that the throughput ratio is larger.
4.3 Performance evaluation of a realistic 802.11e scenario

In the previous section we have studied the functionality of the four variable parameters defined in 802.11e. In the following section we evaluate their effectiveness with a realistic scenario.

4.3.1 Models

Before the performance analysis, we will first present the system and traffic models.

System model
In this part of the study we consider a single BSS operating in the 802.11e mode. In this BSS a number of wireless stations contend for the shared WLAN radio access medium. This WLAN operates with an access rate of 11 Mbps, as can be found in the most deployed WLANs today. Long physical headers and preamble (see section 2.4.2) are applied, to ensure compatibility with 802.11b. We only consider 802.11e HCF/EDCA in its basic mode, since we have previously shown that RTS/CTS mode at high WLAN data rate is not efficient compared to basic access mode. We do not take 802.11e HCF mode into consideration as it is expected that the first generation 802.11e products will only allow HCF/EDCA mode. All four variable parameters specified in HCF/EDCA, including AIFS, CWmin, CWmax and TXOPLimit, are utilized in order to achieve the desired service quality.

Traffic model
The considered WLAN serves several types of users:

• **VoIP**
  In this type of service relatively small packets are sent over the WLAN medium at a fixed packet rate. The traffic is in both the upstream and the downstream direction. The call arrives according to a Poisson process and their duration is taken to have a deterministic value.

  The QoS of this type of service is defined in terms of packet delay, delay jitter and packet loss.

• **Video streaming, VoD**
  Traffic from this type of users is generated in the downstream direction. It contains UDP packets in constant rate with a given size. We consider On/Off sources with exponentially distributed on and off time. Traffic load is expressed in term of average number of active flows, which follows from the total number of sources, average on time and average off time.

  The QoS measures for video streaming include delay jitter and packet loss rate. The priority of this type of traffic will be intermediate, which means lower than VoIP but higher than best effort type of traffic.

• **Web browsing, best effort traffic**
  Traffic data is transmitted in the downstream direction. Because the traffic here consists of TCP streams, there are also TCP ACK packets sent in the upstream direction. The files arrive according to a Poisson process. These files are chosen to have exponential distributed file size with file size mean 1/u. The files are segmented
into packets of a given size (with a final packet containing the flow’s remainder), which are processed at the WLAN’s MAC layer.

We assume for these stations that they use a version of TCP that includes delayed acknowledgement (refer to section 2). The web server, the originator of data, is located distantly from the BSS and connected to the AP by “Internet”, see Figure 3-7. This “Internet” link is assumed to have no loss and a time invariant delay.

This type of traffic has the lowest priority in our scenario. The most important QoS requirement here is in terms of file transfer time.

4.3.2 Simulation results

A number of simulation results are presented below, which are obtained by means of dynamic simulations. The simulations are performed using NS-2 2.26 with TKN HCF/EDCA extension. These simulations include all three type of services described above. The specific parameters of these services are summarized in Table 4-3.

<table>
<thead>
<tr>
<th>Voice over IP</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stream type</td>
<td>UDP CBR</td>
</tr>
<tr>
<td>Direction</td>
<td>Downstream &amp; upstream</td>
</tr>
<tr>
<td>Packet size</td>
<td>200 bytes (IP segment)</td>
</tr>
<tr>
<td>Packet rate</td>
<td>50 packets per second per direction</td>
</tr>
<tr>
<td>Bit rate</td>
<td>80 Kbits per second per direction</td>
</tr>
<tr>
<td>Duration</td>
<td>180 seconds</td>
</tr>
<tr>
<td>Arrival process</td>
<td>Poisson</td>
</tr>
<tr>
<td>Load variable</td>
<td>Average call arrival rate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Video on Demand</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stream type</td>
<td>UDP CBR</td>
</tr>
<tr>
<td>Direction</td>
<td>Downstream</td>
</tr>
<tr>
<td>Packet size</td>
<td>1500 bytes (IP segment)</td>
</tr>
<tr>
<td>Packet rate</td>
<td>40 packets per second</td>
</tr>
<tr>
<td>Bit rate</td>
<td>480 Kbits per second</td>
</tr>
<tr>
<td>Arrival process</td>
<td>OnOff sources</td>
</tr>
<tr>
<td>On/Off time ratio</td>
<td>4:1</td>
</tr>
<tr>
<td>On time distribution type</td>
<td>Exponential</td>
</tr>
<tr>
<td>Off time distribution type</td>
<td>Exponential</td>
</tr>
<tr>
<td>On time average</td>
<td>300 seconds</td>
</tr>
<tr>
<td>Load variable</td>
<td>Total number of source stations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Web Browsing</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stream type</td>
<td>TCP</td>
</tr>
<tr>
<td>Direction</td>
<td>Downstream</td>
</tr>
<tr>
<td>Packet size</td>
<td>1500 bytes (IP segment)</td>
</tr>
<tr>
<td>File size distribution type</td>
<td>Exponential</td>
</tr>
<tr>
<td>File size average</td>
<td>120 Kbits</td>
</tr>
<tr>
<td>Arrival process</td>
<td>Poisson</td>
</tr>
<tr>
<td>Load variable</td>
<td>Average file arrival rate</td>
</tr>
</tbody>
</table>

Table 4-3 Parameter values of the services
For each service the load variable is given in Table 4-3. The WLAN is operating at an access rate of 11 Mbps and in its basic access mode. The important parameter values of the WLAN, based on 802.11b DSSS PHY, can be found in Table 2-3.

Furthermore the total system load is defined as:

$$\rho = \frac{r_{VoIP} + r_{VoD} + r_{WB}}{r_{WLAN}}.$$  

If the load is increased in one type of the services, the total system load will also increase accordingly. In each simulation all three types of services are present. While the total system load increases in each simulation only in one of the services types, the other services have a constant share of system load. Hence, in all simulations, two of the three values are constant. The amount of the constant traffic is shown in Table 4-4. By applying Little’s Law the intuitive characteristics of VoIP and WB can expressed in the load variables in Table 4-3. Given the On/Off time ratio of VoD, the load variable value of the service VoD can also be derived. These values are also included in Table 4-4.

<table>
<thead>
<tr>
<th>Services</th>
<th>parameters</th>
<th>values</th>
</tr>
</thead>
<tbody>
<tr>
<td>VoIP</td>
<td>average number of concurrent calls</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>average call arrival rate</td>
<td>0.01667 calls per second</td>
</tr>
<tr>
<td>VoD</td>
<td>average number of streams</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>total number of source stations</td>
<td>10</td>
</tr>
<tr>
<td>WB</td>
<td>average download date rate</td>
<td>960 Kbps</td>
</tr>
<tr>
<td></td>
<td>WB average file arrival rate</td>
<td>(8 files per second)</td>
</tr>
</tbody>
</table>

Table 4-4 Service characteristics

Our main goal is to determine if 802.11e provides better quality of services for time-stringent services like VoIP and VoD. Therefore the performances of both 802.11b and 802.11e are compared. In all scenarios a maximum of 50 users is imposed to the system.

4.3.2.1 802.11b simulations, system load increased in VoIP traffic.

The performance of an 802.11b WLAN is examined first. The load of the total system is increased with the VoIP call arrival rate. The performance of VoIP traffic is shown in Figure 4-12.
The performance parameters packet delay, delay jitter \(^{20}\) and packet loss ratio are found to downgrade rapidly if the system load is increased in the VoIP type of traffic. As shown in Table 2-2, the packet delay from an end-to-end VoIP call shall not exceed 150ms. The delay at the WLAN, which serves as an access link, shall not exceed 40ms to deliver a reasonable connection. Table 2-2 also shows that the desired packet loss ratio for VoIP traffic is lower than 1%. The upstream performances are within these thresholds. However, the downstream performance does not meet these requirements already at a system load of as low as 0.21.

The performance at the downstream direction is worse than in the upstream direction. This is because the download stream is originated from the AP and the AP must process multiple VoIP streams and packets from other types of services. While the AP does not obtain a larger share of the medium capacity, it has a much larger amount of data to transmit. Hence the VoIP performance is worse in the downstream direction than the upstream direction. The rapid downgrade of the performance parameters are due to the small packet size in the VoIP service. On the WLAN MAC layer all data packets have an equal amount of overhead, including WLAN MAC/PHY header, MAC ACK packets, interframe spacing intervals and backoff windows. Hence, the WLAN medium is very inefficient in sending small packets.

### 4.3.2.2 802.11e simulations

The same scenario is applied to an 802.11e WLAN with different TC parameters. Due to the strict delay requirements for VoIP and VoD traffic, the time interval between the packet transmissions has to be of limited length. This can be achieved by assigning small CWmin and CWmax values to the TC parameters of these services. Because the priority of VoIP is higher than VoD, we choose both CWmin and CWmax for VoIP to be smaller than for VoD. These values are much lower than the 802.11b parameter values. In our scenario WB traffic has the lowest priority and its CWmin and CWmax are assigned same values as found in an 802.11b network. In section 4.2 it is shown that differentiated AIFS values have a positive impact on the channel efficiency, therefore the AIFS values are also different for these three service types. According to their priority, AIFS of WB has a greater value than VoD, which is on its turn greater than VoIP. The resulting TC parameters for the three services are shown in Table 4-5.

---

\(^{20}\) Refer to section 2.2, jitter is the standard deviation of delay times.
Table 4-5 The 802.11e WLAN TC parameter values for three services.

<table>
<thead>
<tr>
<th></th>
<th>TC0</th>
<th>TC1</th>
<th>TC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic</td>
<td>VoIP Up</td>
<td>VoD</td>
<td>WB</td>
</tr>
<tr>
<td>CWmin</td>
<td>7</td>
<td>15</td>
<td>31</td>
</tr>
<tr>
<td>CWmax</td>
<td>63</td>
<td>255</td>
<td>1023</td>
</tr>
<tr>
<td>AIFSN</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Note that $AIFS_{TC} = SIFS + AIFSN_{TC} \times tTimeSlot$, see section 2.4.3.

Figure 4-13 shows the performance of VoIP in such an 802.11e network and the previous 802.11b network if the total system load is increased with the number of VoIP call arrivals.

By applying traffic differentiation with 802.11e, the performance of the high priority VoIP traffic improves. Especially the downstream traffic shows better PLR, jitter and packet delay. However, the PLR still increases rapidly in the downstream direction and becomes unacceptable around a load of 0.23.

In section 4.3.2.1 it is shown that the downstream VoIP traffic is more sensitive if the system is approaching saturation. Figure 4-13 shows that downstream traffic is still a bottleneck if we have only one TC for both upstream and downstream VoIP traffic. Hence, a separate TC is subsequently defined for the downstream VoIP traffic to provide it a higher priority. This is done by applying a TXOPLimit duration in which a number of VoIP packets can be transmitted. The other advantage of applying TXOPLimit is its positive impact on the channel efficiency. Considering these factors, the following parameter set in Table 4-6 is applied to the renewed 802.11e WLAN, in which TC0 has the highest priority and TC3 the lowest.

Table 4-6 The 802.11e WLAN TC parameter values, with an additional TC for downstream VoIP traffic.

<table>
<thead>
<tr>
<th></th>
<th>TC0</th>
<th>TC1</th>
<th>TC2</th>
<th>TC3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic</td>
<td>VoIP Down</td>
<td>VoIP Up</td>
<td>VoD</td>
<td>WB</td>
</tr>
<tr>
<td>CWmin</td>
<td>7</td>
<td>7</td>
<td>15</td>
<td>31</td>
</tr>
<tr>
<td>CWmax</td>
<td>63</td>
<td>63</td>
<td>255</td>
<td>1023</td>
</tr>
<tr>
<td>TXOPLimit</td>
<td>0.009s</td>
<td>0.003s</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AIFSN</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
Figure 4-14 shows the performance of VoIP in an 802.11e network with and without separate TCs for upstream and downstream VoIP traffic. The total system load is increased with the number of VoIP call arrivals.

Although the delay and delay jitter of the packets in the upstream direction become slightly larger, they are still within the required threshold. In addition, the performance of the packets in the downstream direction has improved significantly. Even at a system load of 0.5, all performance measures are still within the required threshold. Note the decrease of PLR in the downstream direction above a load of 0.4. This is due to the large confidence interval with these simulations. If the time allows more replications, the PLR will increase with the system load.

By applying the given TXOPLimit values to the VoIP streams, a much smaller number of contentions are necessary to transmit the same amount of packets and therefore data compared to the 802.11b scenario, in which each VoIP packet has to contend for medium access. Because of the smaller number of necessary contentions, the overhead added by the WLAN MAC/PHY is also smaller and therefore the system is able to carry a much larger number of VoIP calls.

Furthermore the performance of VoIP is examined by increasing the system load with the other two services. Figure 4-15 shows the packet loss ratio of the VoIP downstream in these scenarios. The reason to only display downstream packet loss ratio is because it is the most critical performance measure in the simulation results shown in Figure 4-12 and Figure 4-14.
A minor impact of the lower priority TCs on the performance of VoIP is observed. The VoD TC has a slightly larger impact on the performance of VoIP than WB. This is due to the fact that the TC parameter values of VoD are much closer to the VoIP TC parameter values. Compared to the WB TC, the smaller value of CWmax in VoD TC also has a negative impact on the medium efficiency, which results in less available bandwidth for the services.

The performances of VoD and WB are also examined in the 802.11e network while separately increasing the total system load in all three TCs. Figure 4-16 shows the obtained packet loss ratio for the VoD TC if system load is separately increased in all three TCs.

Figure 4-16 shows that the performance of VoD downgrades rapidly if the system load is increased with the two highest TCs. However, if only more WB traffic is added to the network, the performance of the present VoD streams will not suffer from it. This again shows that 802.11e is able to protect higher priority TCs. An interesting observation is that load increase with VoIP has the least impact on the VoD performance in the very low load region, up to a load of 0.22. A possible explanation is the application of TXOPLimit for the VoIP TC. It increases the channel effective capacity. This has a positive impact on the total system performance. Although the increased load with higher TC will downgrade the performance of VoD, the previously described effect is more dominant in the low loads region.
The file transfer times (page download time) of the WB traffic are shown in Figure 4-17.

![WB File transfer time, 802.11e](image)

**Figure 4-17 The file transfer times of WB by increasing system load with different services**

The load increase with the VoIP traffic results again in the least performance downgrade in the very low load region. However, the page download time increases rapidly if the more VoIP traffic is present such that the total system load is above 0.23. Surprisingly, the performance of WB does not significantly more suffer from load increase with the VoD traffic compared to with WB traffic. Further investigation is required to provide more insight in this observation.

### 4.3.3 Summary

In this section the performance of a network with different types of services are examined, including Voice over IP, Video on Demand and Web Browsing. This is done in both an 802.11b network and an 802.11e network.

Due to the small size of the VoIP packets the system performance downgrades rapidly if more VoIP calls are made concurrently. The 802.11b network is only able to meet the strict requirements with a very small number of concurrent VoIP calls.

By applying prioritized traffic and TXOPLimit in the 802.11e network this problem can be solved. The 802.11e WLAN with the given set of 802.11e parameters is able to provide sufficient quality for the VoIP calls if the load is increased in this type of traffic. Furthermore, the quality of service of VoIP will not suffer from load increase in other service types.

The application of TXOPLimit in the VoIP TCs has a positive impact on the overall system performance if the total system load is low. The increase in load in VoIP in that region has less negative impact on the performances of other services than load increase in these services. However, if the load is high, the VoIP TCs with higher priority will acquire a large amount of the medium capacity. The result is that the quality of other services will downgrade rapidly in this region.
4.4 Conclusion

The first part of this section investigates the impact of the 802.11e parameters CWmin, CWmax, AIFS and TXOPLimit by means of throughput analysis. Furthermore, their impact on the channel efficiency is examined. The results of these investigations can be found in Table 4-1 and Table 4-2. By combining these results, the functions of these parameters can be summarized in Table 4-7.

<table>
<thead>
<tr>
<th>parameters</th>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Protect higher priority TC</td>
</tr>
<tr>
<td>CWmin</td>
<td>-</td>
</tr>
<tr>
<td>CWmax</td>
<td>++</td>
</tr>
<tr>
<td>TXOPLimit</td>
<td>o</td>
</tr>
<tr>
<td>AIFS</td>
<td>++</td>
</tr>
</tbody>
</table>

*Table 4-7 The functions of the 802.11e parameters*

-- highly ineffective/high negative effect; - ineffective/negative effect;
* intermediate; + effective/positive effect; ++ highly effective/high positive effect;*

The performance of an 802.11e WLAN is examined with three types of services, i.e. Voice over IP, Video on Demand and Web Browsing. By applying a given set of 802.11e parameter values, the performance of the VoIP services, which has stringent requirements, is largely improved compared to an 802.11b network. In addition, it is shown that TCs with different parameter sets are able to provide service differentiations, as higher priority TCs are not sensitive to the increase in load in the lower TC services.
5  Concluding remarks

In this research the performance of IEEE 802.11 WLAN is studied. The performance of TCP, with respect to file transfer times, over the vastly deployed IEEE 802.11b is examined. In addition, the IEEE 802.11e, the upcoming WLAN standard with QoS enhancements is investigated. The findings in both parts of the research are summarized below.

5.1  File transfer times with TCP over IEEE 802.11b WLAN

- Simulation with TCP flows of unlimited size delivers inaccurate results towards fairness. In realistic networks TCP flows are of finite size. In these cases the medium capacity is fairly distributed among the flows.

- Two models have been examined to predict the file transfer times: the ordinary Processor Sharing model with a fixed capacity and an integrated packet/flow level model proposed by Lassila et al ([Lassila]). They provide accurate predictions for files of large sizes. However, predictions of both models are generally too optimistic. We have shown that at flow level file transfer with TCP over WLAN behaves alike TCP over a fixed link, if take the capacity of the fixed link as the effective throughput of WLAN at a certain access rate.

- Better TCP models should be able to provide more accurate file transfer time predictions. In future works models may be considered that captures the bursty characteristics of TCP. Other refinements include investigation of achievable rate of small TCP flows, as well as a more detailed research of the impact of slow start on the TCP throughput.

5.2  IEEE 802.11e WLAN

- By differentiating the values in the IEEE 802.11e parameters CWmin, CWmax, AIFS and TXOPLimit, flows can obtain different priorities over the WLAN medium.

- Higher priority can be assigned to a flow by reducing its CWmin, CWmax and AIFS values, and by assigning a larger TXOPLimit. Reduced CWmin and CWmax values have negative impact on the overall channel efficiency while a larger TXOPLimit positively affects the channel efficiency. The latter is also very efficient to improve channel efficiency in networks where some flows consist of small size packets. However, TXOPLimit also introduces more delay with respect to packet transmission and this must be taken consideration in applying IEEE 802.11e network parameters.

- By utilizing differentiated 802.11e parameter values, the high priority traffic classes can effectively be protected in a saturated network. At the same time low priority traffic classes can suffer from starvation effects, especially if prioritized traffic is provided by differentiation through CWmax and AIFS.
• In WLAN’s infrastructure mode, the access point has generally a larger amount of traffic to transmit than other stations. Hence the AP is often the bottleneck in data transmission in a WLAN. The performance can improve if the downstream flows of the same type services are assigned a higher priority than in the upstream direction.

• Further research should include more detailed investigation of the impact of the four parameters on the packet level system performances, in order to assign more accurate parameter values according to the network and service requirements. Furthermore, it is also worthwhile to examine the additional mode defined in IEEE 802.11e: HCF/HCCA and compare its performance with HCF/EDCA.
References


[802.11i] http://www.commsdesign.com/showArticle.jhtml?articleID=22102230


[Aad] I. Aad, C. Castelluccia, Differentiation mechanisms for IEEE 802.11, in proceeding of INFOCOM, 2001


[Botti] M. Bottigliengo, C. Casetti, C.F. Chiasserini, M. Meo, Short-term Fairness for TCP Flows in 802.11b WLANs, in proceeding of INFOCOM, 2004


[ITU-E800] International Telecommunication Union, Terms and definitions to quality of service and network performance including dependability, ITU-T E.800, 1994


[Qiao] D. Qiao and K.G. Shin, Achieving Efficient Channel Utilization and Weighted Fairness for Data Communications in IEEE 802.11 WLAN under the DCF, in proceeding of the Tenth International Workshop on Quality of Service, 2002


<table>
<thead>
<tr>
<th>Reference</th>
<th>Citation</th>
</tr>
</thead>
</table>