

Tor Kjetil Moseng

A QoS Architecture for Mobile Ad Hoc Networks

Thesis for the degree of Philosophiae Doctor

Trondheim, June 2009

Norwegian University of Science and Technology
Faculty of Information Technology,
Mathematics and Electrical Engineering
Department of Telematics



NTNU

Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Information Technology, Mathematics and Electrical Engineering
Department of Telematics

© Tor Kjetil Moseng

ISBN 978-82-471-1604-3 (printed ver.)

ISBN 978-82-471-1605-0 (electronic ver.)

ISSN 1503-8181

Doctoral theses at NTNU, 2009:111

Printed by NTNU-trykk

Abstract

A Mobile Ad Hoc Network (MANET) is a shared wireless network without any infrastructure, consisting of mobile nodes connected by wireless links. The nodes are free to move and organize themselves arbitrarily. The nodes in the network are therefore depending on each other in order to communicate over multiple hops. Due to the physical characteristics of wireless networks, the channel is time-varying, which makes it hard to both predict and sustain a bit rate level. The nodes' mobility causes topology changes, and further load and capacity variations. Traditional usage areas are battlefield and disaster areas, while new areas like extended network coverage and gaming are emerging.

Quality of Service (QoS) is needed in every network in order to differentiate traffic with different performance requirements, e.g. voice and e-mail applications. Providing QoS in wireless environments with varying conditions is complex, and hard guarantees can not be given. Consequently, the aim is to give differentiated treatment to traffic with different performance requirements. In addition, we can not study the MANET without considering fixed networks. Communication with fixed networks is important, for example by accessing the Internet.

In this thesis the Differentiated Services (DiffServ) architecture is applied and adapted to MANETs. Using the same QoS architecture will ease the transition between the wireless and wired domain. But the special characteristics of wireless networks require modifications to the original DiffServ architecture. In investigations there was found restrictions on the number of classes to use, and this number was dependent on the type of traffic in the network. A QoS architecture based on the DiffServ framework is proposed, with an admission control based on the concept of shadow classes, and Explicit Congestion Notification (ECN) to avoid congestion. New flows are tested in a shadow class before getting admission to the network and its designated class. The shadow class has the same scheduling properties as the designated class, but is differentiated by a higher drop probability in the buffers. Both the admission control and ECN are thus build on the same principle by controlling the load from probabilistic functions in the buffers, and are studied to find their individual and combined effects.

In wireless environments the probability of a packet loss increases with the number of hops, which gives services an unpredictable performance for users. A predictable service, independent of number of hops, is provided by scheduling based on the path

information; the packets are differentiated based on the number of hops made or left to make, increasing the predictability at the cost of performance.

Keywords Admission control, congestion control, DiffServ, ECN, fairness, mobile ad hoc networks, QoS, path dependent scheduling, predictability, shadow classes.

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree of philosophiae doctor (PhD) at the Norwegian University of Science and Technology (NTNU).

The work has been carried out in the period August 2004 to February 2009. Besides the research work, it has included compulsory courses corresponding to one semester of full-time studies, a quarter of a year of teaching assistance and student guidance, and two leave of absences (each on 29 weeks) due to the birth of my two sons.

During the study period, I have been hosted and funded by the Centre for Quantifiable Quality of Service in Communication Systems (Q2S), Centre of Excellence. Q2S is funded by the Norwegian Research Council, NTNU and Uninett. The studies were formally conducted at the Department of Telematics, Faculty of Information Technology, Mathematics and Electrical Engineering. Professor Øivind Kure and Professor Peder J. Emstad, both at NTNU, has been the supervisor and co-supervisor, respectively, of this work.

The document has been formatted in L^AT_EX using a modified version of the document class *kapproc.cls* provided by Kluwer Academic Publishers.

Acknowledgements

A number of people have directly or indirectly contributed to the work presented in this thesis. First of all I would like to thank my supervisors Øivind Kure, who also are the co-author of the papers included in this thesis, and Peder J. Emstad. Your contributions have been many and invaluable during the work. Then I would like to thank my office mate Astrid Undheim for all the discussions we have had, both of professional and personal art, and for the motivation you have given me. You have been a great room mate during these years. For helping me out with the L^AT_EX formatting, I would like to thank Anders Mykkeltveit. I would also like to thank all colleagues at Q2S who have created a very open and enjoyable atmosphere - especially in coffee breaks and lunches. A special attention goes to the administrative and technical staff at the Q2S Centre; thanks to Anniken Skotvoll and Mette Veronica Olsen for handling all administrative matters, and to Hans Almåsbygg for providing a reliable working environment.

Finally, I would like to thank my family. I am very grateful for the continuous love and support from my parents and my sister. But most of all, I would like to thank my dear Marit and my two sons, Ådne and Håvard, for always being there for me, and cheering me up in troublesome times during the work on this thesis.

Contents

Abstract	iii
Preface	v
Acknowledgements	vii
Abbreviations	xi
List of Papers	xv
Part I Thesis Introduction	
1 Background	5
2 Quality of Service in IP Networks	15
3 Fairness	31
4 Research Methodology	34
5 Thesis Context	34
6 Contributions	35
7 Conclusions	42
Part II Included Papers	
Paper A: DiffServ in Ad Hoc Networks	49
<i>Tor K Moseng, Oivind Kure</i>	
1 Introduction	49
2 Simulation Setup	51
3 Results	52
4 Aggressive Best-Effort Traffic	54
5 Fairness	56
6 Conclusions	59
References	60
Paper B: QoS Architecture in Ad Hoc Networks	65
<i>Tor K Moseng, Oivind Kure</i>	
1 Introduction	65
2 The Shadow Class	66
3 The QoS Architecture	67
4 Simulation Setup	69
5 Simulation Results	70
6 Conclusions	74
References	74
Paper C: QoS Architecture in Ad Hoc Networks: Effects of Shadow Classes and ECN to Regulate the Load	79

<i>Tor K Moseng, Oivind Kure</i>		
1	Introduction	79
2	Background	80
3	The QoS Architecture	81
4	The ECN Mechanism	83
5	Simulation Setup	84
6	Results	85
7	Conclusions	88
	References	89
Paper D: Evaluation of Path Dependent Scheduling in Ad Hoc Networks: a Suitable Fairness Mechanism?		93
<i>Tor K Moseng, Oivind Kure</i>		
1	Introduction	93
2	Related Work	95
3	The QoS Architecture	96
4	Fairness	97
5	Path Dependent Scheduling	97
6	Simulation Scenario 1: Simple Chain Topology	98
7	Simulation Scenario 2: Random Topologies	101
8	Conclusions	105
	References	105
Part III Thesis Appendix		
Paper E: QoS Architecture in Ad Hoc Networks: Sensitivity Analysis of the Admission Control		109
<i>Tor K Moseng, Oivind Kure</i>		
1	Introduction	109
2	Related Work	110
3	Sensitivity Settings	111
4	Sensitivity Analysis	113
5	Conclusions	116
	References	116
Bibliography		119

Abbreviations

AC	Access Category
ACK	Acknowledgement
AF	Assured Forwarding
AIFS	Arbitrarily IFS
AIMD	Additive Increase Multiplicative Decrease
AODV	Ad Hoc On-demand Distance Vector
AQM	Active Queue Management
aRSVP	Aggregated RSVP
ARQ	Automatic Repeat-Request
ASAP	Adaptive Reservation and Pre-allocation Protocol
BA	Behavior Aggregate
BE	Best-Effort
CAC	Call Admission Control
CBR	Constant Bit-Rate
CCID	Congestion Control Identifier
CEDAR	Core-Extraction Distributed Ad Hoc Routing
CLAD	Cross-layer Architecture for DiffServ
CoV	Coefficient of Variation
CSMA/CA	Carrier-Sense Multiple Access with Collision Avoidance
CTS	Clear To Send
CW	Collision Window

CWND	Congestion Window
DARPA	Defense Advanced Research Projects Agency
DCCP	Datagram Congestion Control Protocol
DCF	Distributed Coordination Function
DiffServ	Differentiated Services
DIFS	DCF IFS
DS-SWAN	Differentiated Services - SWAN
DSCP	Differentiated Services Code Point
DSDV	Destination-Sequenced Distance-Vector
DSR	Dynamic Source Routing
DYMO	Dynamic MANET On-demand
ECN	Explicit Congestion Notification
EDCA	Enhanced Distributed Channel Access
EF	Expedited Forwarding
EIFS	Extended IFS
EMBAC	End-to-End Measurement-Based Admission Control
FCC	Federal Communications Commission
FQMM	Flexible Quality of Service Model for Mobile Ad Hoc Networks
GPS	Global Positioning System
HCCA	HCF Controlled Channel Access
HCF	Hybrid Coordination Function
IEEE	Institute of Electrical and Electronics Engineers, Inc.
IETF	Internet Engineering Task Force
IFS	Interframe Space
IntServ	Integrated Services
IP	Internet Protocol
ITU-T	Telecommunication Standardization Sector of the International Telecommunication Union

JFI	Jain's Fairness Index
MAC	Medium Access Control
MANET	Mobile Ad Hoc Network
MIB	Management Information Base
MPR	Multipoint Relay
MQRD	Multipath QoS Routing Protocol of supporting DiffServ
NAV	Network Allocation Vector
OLSR	Optimized Link State Routing
PCF	Point Coordination Function
PHB	Per-Hop Behavior
QoS	Quality of Service
RED	Random Early Detection
RFC	Request For Comment
RMP	Reactive MANET Protocol
RSVP	Resource Reservation Protocol
RTS	Request To Send
SD-AODV	Service Differentiation - AODV
SIFS	Short IFS
SLA	Service Level Agreement
SNR	Signal-to-Noise Ratio
<i>ssthr</i>	Slow Start Threshold
SWAN	Service Differentiation in Stateless Wireless Ad Hoc Networks
TCA	Traffic Conditioning Agreement
TCP	Transmission Control Protocol
TE	Traffic Engineering
TEQUILA	Traffic Engineering for Quality of Service in the Internet, at Large Scale
TFRC	TCP-Friendly Rate Control

TORA	Temporally Ordered Routing Algorithm
Tspec	Traffic Specification
TXOP	Transmission Opportunity
UDP	User Datagram Protocol
UP	User Priority
VoIP	Voice over IP
WLAN	Wireless Local Area Network
WMN	Wireless Mesh Network
ZRP	Zone Routing Protocol

List of Papers

Publications Included in the Thesis

- Paper A:
Tor K. Moseng, Øivind Kure. *DiffServ in Ad Hoc Networks*. Presented at the EuroNGI Wireless and Mobility workshop. Sitges, Spain, June 7-9, 2006. Article in Lecture Notes in Computer Science (LNCS). Vol. 4396/2007, Springer Berlin / Heidelberg, May, 2007.
- Paper B:
Tor K. Moseng, Øivind Kure. *QoS in Ad Hoc Networks*. Proceedings of the 2nd International Conference on Communications and Networking in China (ChinaCom). Shanghai, China, August 22-24, 2007.
- Paper C:
Tor K. Moseng, Øivind Kure. *QoS Architecture in Ad Hoc Networks: Effects of Shadow Classes and ECN to Regulate the Load*. Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC). Las Vegas, Nevada, USA, March 31 - April 3, 2008.
- Paper D:
Tor K. Moseng, Øivind Kure. *Evaluation of Path Dependent Scheduling in Ad Hoc Networks: a Suitable Fairness Mechanism?* Proceedings of the Sixth International Conference on Wireless On-demand Network Systems and Services (WONS). Snowbird, Utah, USA, February 2-4, 2009.
- Paper E:
Tor K. Moseng, Øivind Kure. *QoS Architecture in Ad Hoc Networks: Sensitivity Analysis of the Admission Control*. Presented at the EuroNF Wireless and Mobility workshop. Stockholm, Sweden, October 23-24, 2008.

Papers A-D are included in Part II of this thesis, while Paper E is included in Part III to support Papers B-D. Note that some of the published papers have been subject to minor editorial changes before their inclusion in this thesis.

I

THESIS INTRODUCTION

Introduction

An ad hoc network is an autonomous collection of nodes that communicate without any infrastructure or centralized administration over a shared wireless channel. The communication is directly between two or more nodes. If the distance between the communicating parties is longer than the transmission range, the information must be forwarded by intermediate nodes, which act as routers (i.e. multi-hop communication). Figure 1 depicts a typical ad hoc network where the links illustrate a node's next hop neighbors. The nodes may be mobile, which in case changes the topology continuously. The network is then termed as a Mobile Ad Hoc Network (MANET). The term *ad hoc* means that the grouping of nodes happens ad hoc by a need for communication.

One of the first multi-hop packet radio networks was made by the Defense Advanced Research Projects Agency (DARPA) in 1973 for military operations [Kah77, JT87]. It was made for highly mobile scenarios in defense and battlefield environments, where the existing infrastructure, if any, should be avoided. Other usage scenarios are disaster areas and emergency situations, where the infrastructure may not be available or has been destroyed. Today, other usage scenarios have emerged; one example is to increase the coverage area of fixed access points like a Wireless Local Area Network (WLAN) hot-spot, while another example is gaming between random participants in a café or a bus. The papers by Corson et al. [CMC99] and Frodigh et al. [FJL00] give more introductory reading on MANETs.

The usage scenarios depicted will demand support for real-time services, and these services will have Quality of Service (QoS) requirements for the delivered performance. The requirements can typically be stated by parameters like loss, delay, jitter, and bandwidth. Ensuring a high quality on the services received, requires QoS provisioning end-to-end in the network. Also, the operations in scenarios like battlefields, emergency, or disaster areas, demand a degree of predictability. Providing a predictable delivery of QoS in MANETs is challenging due to the characteristics of wireless networks, and especially with the presence of mobility. First of all, the available capacity in a wireless network is more constrained than its wired counterpart since the initial capacity is lower, and a node's transmissions will affect all nodes within range. This necessitates that the offered traffic is subject to admission control for avoiding degraded service of the existing traffic. Second, the topology may change quite rapidly due to mobility, causing variability in capacity and traffic load patterns. Consequently, congestion control is therefore needed.

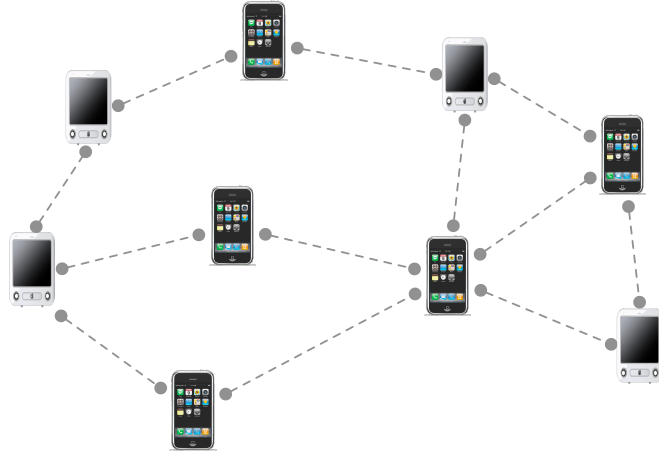


Figure 1. A Mobile Ad Hoc Network (MANET)

The need for QoS support in MANETs has often been solved without regarding the fixed network. Communication with fixed network nodes must then be relayed through gateways with the risk of precision loss in the QoS mapping. The initial motivation for the work in this thesis was to extend the Differentiated Services (DiffServ) [BBC⁺98] architecture into MANETs. Having one common QoS architecture independent of network technology would ease the integration of MANETs into fixed networks, and improve the end-to-end QoS support. The DiffServ architecture has some important properties that make it able to support QoS end-to-end over different mediums. First of all, it scales well. Second, different policies can be defined for different domains, which make DiffServ able to handle the requirements in both a wireless and wired domain. And third, differentiating the service classes without any hard guarantees complies better with the characteristics of wireless networks. Also, it is important to look at solutions for the fixed network since they introduce well proven Traffic Engineering (TE) techniques, which have been adopted by proposals for QoS in MANETs.

To protect the existing traffic from degraded service, an admission control based on the concept of shadow classes is proposed in this thesis. It is a source-based admission control, which induce low overhead and complexity. A best-effort Medium Access Control (MAC) layer is used in the architecture. This means that packets from different nodes can not be differentiated on the MAC layer; the same parameters are used for accessing the channel independently of priority. Consequently, a flow may affect traffic in its designated class if scheduled from a lower prioritized class. Therefore, preemption is used when the network can not accommodate the new flow. For avoiding congestion due to a dynamic topology, Explicit Congestion Notification (ECN) [RFB01] is used to notify sources that may preempt their ongoing sessions.

Routing is important for the connectivity in multi-hop networks. The routing process can include QoS requirements that must be met by all intermediate nodes for end-to-end QoS support (i.e. *QoS routing*). QoS routing can improve the flow blocking ratio and the resource utilization, and could be an alternative to the admission process in this thesis. But QoS routing would still need source admission and preemption in congestion situations. In this thesis the proposed QoS architecture is decoupled from the routing process, avoiding admission control at every hop along the path.

The predictability of a service's performance is addressed by using the path information in the packet scheduling. Resources are allocated for more equal conditions independent of path length.

The main part of this thesis, Part II, is a collection of four papers (Papers A-D) analyzing different aspects of the QoS architecture in this thesis. In addition, Part III includes a paper (Paper E) to support Papers B-D in Part II. The rest of Part I consists of background information and relevant research, along with contributions from this thesis. Section 1 gives necessary background information on MANETs, where Carrier-Sense Multiple Access with Collision Avoidance (CSMA/CA), routing, and capacity issues are emphasized. Section 2 considers QoS approaches for both fixed networks and MANETs. In addition, QoS approaches for MANETs interconnected with fixed networks are considered. Section 3 discusses the fairness, and emphasizes the importance of predictability. Section 4 defines the research methodology for this thesis. Section 5 describes the thesis context; it states this work in relation to the DiffServ architecture. Section 6 gives the contributions for each paper, before the work is concluded in Section 7.

1. Background

This section discusses state of the art in issues related to the thesis. Section 1.1 lists special characteristics and challenges in general ad hoc networks. One of these characteristics is distributed operation. In this thesis the CSMA/CA is used, and it is described to provide a basic understanding of the operations involved for a best-effort channel access, which give limitations for the IP layer. The routing is not directly targeted by the work in Part II, but it is important to have a basic understanding of the classification of different routing approaches - especially related to QoS. Section 1.2 considers the routing. Section 1.3 considers the capacity in ad hoc networks, and how it may vary in dynamic environments. Wireless Mesh Networks (WMNs) provide a wireless infrastructure for communication and Internet access. Ad hoc networks can be regarded as one type of WMNs. Section 1.4 briefly presents WMNs since they may co-exist with ad hoc networks for improved connectivity and capacity.

1.1 Mobile Ad Hoc Networks

A MANET is, as mentioned in the introduction, an autonomous collection of nodes that communicate without any infrastructure or centralized administration over a shared wireless channel. The nodes are in addition free to move, which gives a dynamic topology. MANETs are very different from traditional fixed networks, with

characteristics that need special attention. Some characteristics related to MANETs are listed below [Gei02].

- **Distributed operation**
MANETs are usually self-organizing and self-configuring, operating in a decentralized manner without any infrastructure. This situation necessitates efficient solutions for functions like security and routing.
- **Dynamic topology**
The nodes are mobile, which will in particular challenge the routing protocols to maintain the network's connectivity. In addition, the surrounding conditions can make a link unavailable, or a power outage can make a node unavailable.
- **Varying capacity**
There will always be a probability of bit-errors over a wireless channel. However, this probability is time-varying, and will result in a time-varying link capacity. In addition, changes in the topology will change the traffic pattern and the interference, leading to variations in the available capacity.
- **Low-power devices**
A mobile node will in most cases be powered by a battery with a limited lifetime. Nodes out of power will change the topology and cause packet losses and re-routings. Not only will this affect the network connectivity, but also the information processing, memory usage, and transceiver functions will be affected if nodes must be cautious in their power consumption. A node's remaining power can be a parameter in the routing process.
- **Hidden and exposed nodes**
Hidden nodes may cause collisions because a hidden node senses the channel to be idle when it is not. Exposed nodes may lead to underutilized capacity because of dropped sessions that could have been active in parallel.

The history of MANETs dates back to the work on the ALOHA network, initiated in 1968 [Kuo74]. ALOHA was a single-hop network with distributed channel access management. Based on this work, DARPA made a multi-hop packet radio network in 1973 for military operations [Kah77, JT87]. In 1985, the Federal Communications Commission (FCC) made the Industrial, Scientific and Medical (ISM) bands available for use without any license [Fed], which promoted developments for wireless networking [Gei02]. Tobagi provides an early modeling and performance analysis of multi-hop packet radio networks [Tob87].

The Institute of Electrical and Electronics Engineers, Inc. (IEEE) began in the late 1980's the development of WLANs, and the IEEE 802.11 standard [IEE07] was approved in 1997. IEEE 802.11 provides distributed channel access mechanisms to allow different devices to co-exist within the same area, and has ever since been the most widespread and used technology for WLANs. Different amendments have been developed from the IEEE 802.11 standard. These amendments differ in the modulation scheme and operating frequency, which lead to different physical data rates. Today,

most wireless mobile devices, and access points, comply with either the IEEE 802.11b or IEEE 802.11g standards [Gei04], which have a physical data rate up to 11 Mbps and 54 Mbps, respectively.

Two closely related types of networks are WMNs and sensor networks. WMNs are more or less static ad hoc networks, which can operate as a wireless infrastructure for mobile clients. The topic is covered in Section 1.4. Sensor networks consist of communicating low-power monitoring devices, which are typically assigned only a few tasks like monitoring environmental conditions or registering movement in an alarm system. Activities in sensor networks must have a low consumption of factors like power, processing and memory, and require often very high scalability. Akyildiz et al. [ASSC02] provide a good overview of issues as the protocol stack and network design for sensor networks. From recent advances in the field, a low cost version of the Transmission Control Protocol (TCP) over IP (TCP/IP), which is IPv6 compliant, has been deployed through the Contiki project [Thea] to allow for global identification and accessibility for all sensor nodes.

1.1.1 The Medium Access Control Layer

The MAC layer is a part of the Link layer in the protocol stack. Its main objectives are to access the wireless channel, and to transmit and receive data in contention with other nodes on the same channel. In addition, the MAC layer provides authentication and privacy functions [Gei02]. However, this section focuses on the wireless channel access, and does not consider the operations involved with joining a network or the authentication process.

The IEEE 802.11 standard defines two modes of operation, one for infrastructure-based networking and one for ad hoc networking. The functions are the Point Coordination Function (PCF) and the Distributed Coordination Function (DCF), respectively. In PCF, the access point controls the transmissions from the stations within the range by a point controller to give a contention-free access. On the other hand, DCF is contention-based and distributed in operation, which makes it suitable for MANETs. It is described in more detail below in this section.

While suitable for general ad hoc networking, the IEEE 802.11 does not provide QoS. Due to the shared channel and equal channel access for all packets, a high priority packet at one node will be affected by lower prioritized packets at other nodes. Implications can be higher delay and loss probability for the high priority packets. The MAC layer will thus give limitations for the QoS mechanisms at the IP layer. It is important to have this in mind when analyzing QoS solutions at the IP layer if the best-effort IEEE 802.11 is used. On the other hand, only QoS extensions at the MAC layer are not sufficient for end-to-end QoS in dynamic environments like MANETs; admission control and preemption are still needed. One MAC layer implementation with QoS support is the IEEE 802.11e standard. It introduces a new distributed access function, which is basically a QoS extension to the DCF. It is described in more detail among the QoS solutions for MANETs in Section 2.4.

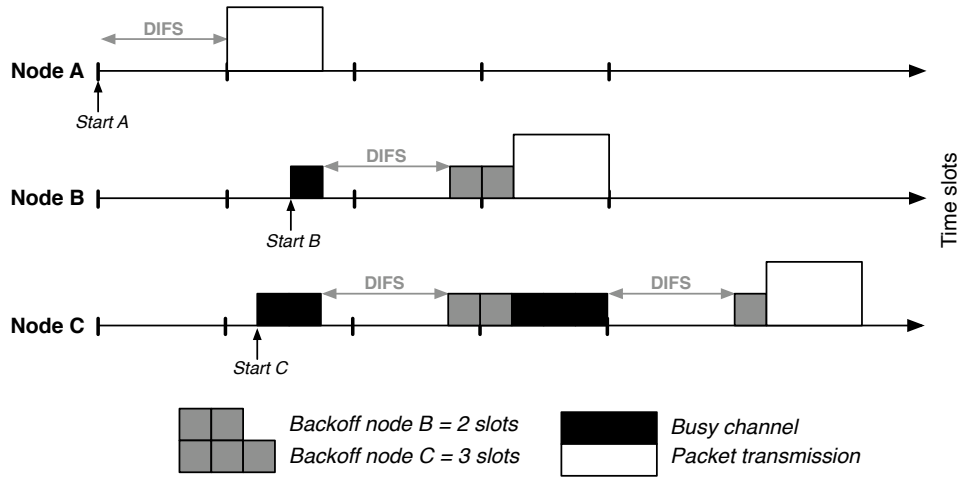


Figure 2. The basic CSMA/CA layer operation

The Distributed Coordination Function

The MAC layer must sense whether the channel is busy or idle before a transmission to avoid collisions. In DCF, the CSMA/CA is used. The CSMA/CA protocol provides both a physical and a virtual sensing mechanism. The actual physical sensing is done by the Physical Layer, but the MAC layer handles the information and acts accordingly. Different time intervals, i.e. Interframe Spaces (IFSs), have been specified to provide different priority levels on the channel access when sensing an idle channel. Under DCF there are three IFSs (listed with increasing duration): Short IFS (SIFS), DCF IFS (DIFS), and Extended IFS (EIFS). Let a simple example illustrate the basic operation of the CSMA/CA; Figure 2 shows three nodes that share the channel without any collisions. When node A wants to transmit, the channel must be idle for a duration greater than the DIFS. Nodes B and C start to sense the channel during node A's transmission, and have to wait a DIFS after the packet transmission. In addition, each node sensing a busy channel must wait some extra slots after the DIFS. The probability of collisions is higher immediately after a packet transmission since there might be several nodes that defer their transmissions during an existing transmission, and wants to transmit as soon as the channel becomes idle. Therefore each node draws a random backoff interval between $[0, CW]$ slots, where CW is the collision window specified between a minimum and maximum value found in the MAC layer's Management Information Base (MIB). Node B draws the shortest backoff interval in the figure, and transmits before node C. Notice that the use of Acknowledgements (ACKs) are omitted for simplicity.

To confirm a packet transmission, the receiver waits a SIFS before it sends an ACK. Since a SIFS is shorter than a DIFS, the ACK will get priority over new data packets.

The sender can not distinguish between a collision and a transmission error, so if an ACK is not returned to the sender, the sender retransmits the packet after a DIFS and a new backoff interval. The *CW* increases exponentially for each transmission, increasing the backoff duration, causing each sender to reduce its load. To control the number of retransmissions, a short and long retry limit is used. How many retries to use will depend on the packet size. All these parameters are found in the MIB. The retransmit process is commonly referred to as Automatic Repeat-Request (ARQ).

In parallel to the physical sensing, each node performs virtual sensing of the channel by listening to the duration field of the MAC frames. The duration is placed in the nodes' Network Allocation Vector (NAV). The NAV will therefore give other nodes' impending use of the channel. An intention to transmit can also be announced through the optional reservation messages Request To Send (RTS) and Clear To Send (CTS). A source may transmit an RTS specifying that it wants to transmit a packet. The receiver will then reply with a CTS. Upon hearing one of these messages, a node will update its NAV accordingly, and refrain from any transmissions. The RTS/CTS handshake avoids some of the hidden terminal problems by preventing surrounding terminals from transmitting at the same time. However, RTS/CTS gives a considerable overhead when preceding small data packets such as voice packets. A threshold (i.e. the *RTSThreshold*) can be specified in order to use RTS/CTS only for large data packets, which are more prone to errors compared to small data packets.

1.2 Routing

One of the major research areas within MANETs is routing strategies. Challenges in MANET routing include mobility, overhead, scalability, security mechanisms, Internet gateway discovery, and QoS support [RT99, RG03]. QoS routing is an important field within the research area, and standardization has been initiated by the Internet Engineering Task Force (IETF) [IET]. Two QoS routing schemes are described in Section 2.4.

According to the routing protocols' operation, they can be classified into four categories: proactive routing, reactive routing, hierarchical routing, and geographical routing. These categories are briefly reviewed below, before the routing is discussed in relation to the focus of this thesis.

1.2.1 Proactive Routing

In proactive routing protocols, also called *table-driven*, every node maintains continuously routes to all other nodes in the network. This necessitates periodically flooding of routing information throughout the network. The Destination-Sequenced Distance-Vector (DSDV) [PB94] was an early available distance-vector protocol. Here, each source has a routing table entry for every destination along with a distance and freshness stamp (a sequenced number). Optimized Link State Routing (OLSR) [CJ03] is an example of a link-state protocol. In OLSR, every node keeps a map of the network in the form of a directed graph, and uses this information for next hop

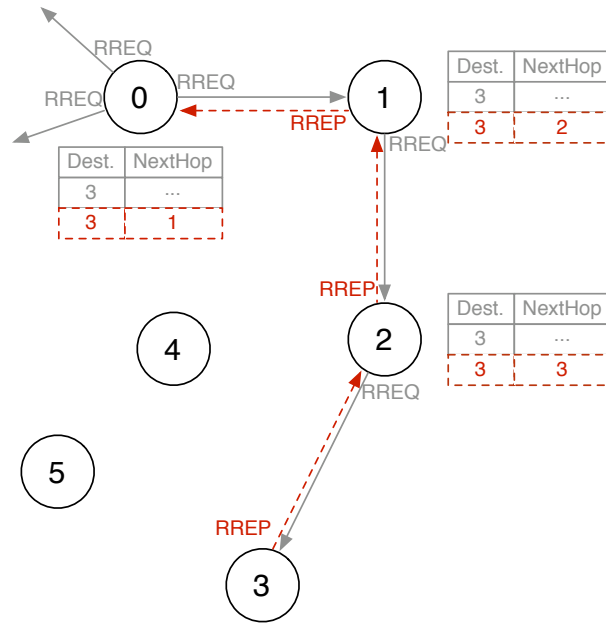


Figure 3. The Ad Hoc On-demand Distance Vector (AODV) routing process

calculations. To limit the flooding, only nodes chosen as Multipoint Relays (MPRs) can exchange link-state information, which also reduces the size of the routing packets.

The advantage with proactive routing is that paths are readily available when needed. However, the disadvantages are a considerable overhead in certain network scenarios (e.g. for low mobility), and the requirements are often higher regarding memory and processing power compared to reactive approaches.

1.2.2 Reactive Routing

In reactive routing protocols, also called *source-initiated* or *on-demand*, a route is discovered when needed, and it is only maintained for the session duration. Examples of reactive routing protocols are Dynamic Source Routing (DSR) [Joh94, JM96] and Ad Hoc On-demand Distance Vector (AODV) [PR99, PBRD03]. In DSR, a source includes addresses of all intermediate nodes to use in a packet. These addresses are obtained by sending route discovery packets from source to destination. In AODV, on the other hand, only the next hop to a destination is stored in a node's routing table. These table entries contain the distance and freshness stamp (from the approach of DSDV). The information is obtained through flooding Route Requests (RREQs) from the source, and a Route Reply (RREP) sent along the reversed path of the RREQ. Each intermediate node updates its routing table with new information (denoted by the sequence numbers). Figure 3 shows the routing process in AODV, where node 0 requests a path to node 3. The reverse path is maintained by keeping a record of the

neighbor from which it first received the RREQ, and keeping this for enough time to answer with a RREP.

The main advantage of reactive routing is less overhead, at least in low mobility scenarios; control messages are only sent when needed. A disadvantage is the delay before the route is established. A route is not available before a path is found, and in worst case RREQs are flooded to the whole network. Also, since a route is only maintained by those nodes needing it, it may take some time to repair a broken connection.

In an effort to standardize IP routing functionality, the IETF MANET Working Group tries to define a Reactive MANET Protocol (RMP) [IET]. During this work, they have made an Internet-Draft on a reactive routing protocol, which builds on AODV, in particular, and DSR. The draft is the Dynamic MANET On-demand (DYMO) routing protocol [CP08]. DYMO has basically the same functionality as AODV, but is easier to implement and better designed for future enhancements.

1.2.3 Hierarchical Routing

Hierarchical routing tries to solve the scalability problems in MANETs. The main idea is to divide the routing into different hierarchical levels, and utilize different routing approaches between the levels. Hybrid routing, which is hierarchical in nature, combines proactive and reactive routing in order to extract the best out of both types. The Zone Routing Protocol (ZRP) is an example of the hybrid approach [HP01]. ZRP defines a zone around each node consisting of its k -neighborhood, where k is the distance in hops. Proactive routing is done within the zone (i.e. nodes no more than k hops away), while reactive routing is done between the zones by requesting their border nodes (i.e. nodes k hops away) for the destination. ZRP defines one protocol for each routing mode.

The advantage with hierarchical routing is that frequently used routes at one level can be proactive to minimize delay, while routes to more distant nodes are created on-demand. However, overhead in the selection of nodes for the reactive routing, and mis-designed hierarchical levels may lead to low efficiency. The best mix of routing strategy will clearly depend on the scenario and network conditions.

1.2.4 Geographical Routing

Geographical routing means to route packets from the source's position relative to its one-hop neighbors' and the destination's positions [MWH01]. Geographical routing consists of two parts: the location service and the routing strategy. The location services can be classified by the number of nodes participating and the completeness of their location databases. The services are termed *some-for-some*, *some-for-all*, *all-for-some*, and *all-for-all*, where for example *some-for-all* means that some of the network nodes are participating, while their databases contain all nodes' locations. There will thus be a trade-off between availability and overhead. The three main routing strategies are greedy forwarding, restricted directional flooding, and hierarchical approaches. Different geographical routing approaches are described in a survey by Mauve et al.

[MWH01]. Today, many wireless devices have the Global Positioning System (GPS) [Theb] implemented, which can be used to find nodes' locations.

1.2.5 Discussion

The proposed QoS architecture is decoupled from the routing protocol; no specific class of routing protocols is necessary for the architecture to function. The architecture ensures the QoS on the paths selected by a routing protocol. However, this does not mean that the performance will be independent of the protocol used. Depending on the traffic pattern, node density and node movement, the protocols have their separate advantages. The scenarios used in the analysis are characterized by low mobility with reasonable in-frequent changes in traffic pattern. Also, the node density is fairly low. For such scenarios an on-demand protocol provides low overhead. In addition, an available implementation of the AODV routing protocol was readily available. It was therefore used as the routing protocol in the simulations.

The simulations used the standard procedure of assigning the highest priority to the routing messages. It minimizes the impact traffic load will have on the performance of the routing protocol. The overhead of the routing protocol is low, and the overall impact of the practice is small on the results. However, the simulations are run without a QoS aware MAC layer causing delay on the routing packets. In addition, there will always be losses due to interference.

A routing protocol maximizes an object function. For routing protocols in fixed networks, the typical object function is to minimize the cost along a path. For ad hoc networks the object function is typically to minimize the number of hops. But the routing process can also be coupled with resource allocation for QoS guarantees through the network. QoS routing is described in more detail in Section 2.4. However, the argument of decoupling of the QoS mechanisms and the routing is still partially valid as long as the routing process operates on a larger time granularity than the lifetime of flows. Call admission and preemption will still be needed. If the routing process is designed to do allocation per flow, call admission and preemption will be part of the routing process. QoS routing is then an alternative.

1.3 Capacity in Wireless Networks

The capacity in wireless networks is dependent on the specific modulation technique used, which gives the raw data rate at the Physical Layer. However, a node's available capacity is not only limited by the raw data rate. Other transmitting nodes in the network imply interference, which lowers the available capacity. Every node must keep quiet within a transmitting node's transmission range to avoid disrupting any data. But also within the wider carrier-sense range, which is typically twice the transmission range, nodes will be affected by a transmitting node. In addition, the capacity can vary based on the surroundings and environmental conditions.

In multi-hop MANETs, the connectivity depends on wide transmission radii, or packet relaying. Decreasing a node's coverage area will initially decrease the interference. But, reducing the coverage area lowers the network's connectivity, and will

consequently increase the number of hops needed between source and destination. And more hops for the same connection add additional load to the network. The relation between the transmission range and the capacity was studied by Kleinrock and Silvester [KS78], in where the authors find the network throughput as a function of the node degree. This function gives a trade-off between many hops and wide transmission radii. The analysis by Gupta and Kumar [GK99, GK00] found that the upper bound on a node's throughput in a random network is $\Theta(W/\sqrt{n \log n})$, based on the authors' protocol model with every node capable of transmitting W bps. With optimal scheduling and node placement, the upper bound reaches $\Theta(W/\sqrt{n})$. Thus, the interference becomes quite severe in dense networks.

Communicating over a multi-hop path results in a considerable reduction in the effective capacity. One problem is that contention resolving is needed at each hop. Another problem is that successive packets belonging to a single connection interfere with each other as they are transmitted along the multi-hop path. Consider Figure 4 which shows a chain of nodes from node 1 (source) to node 6 (destination), where the nodes are placed 200m apart with a transmission range of 250m (solid line) and carrier-sense range (dotted line) of 550m. Successive packets are sent down the chain, and when node 1 transmits to node 2, node 3 must defer its transmission. But, also node 4 must defer its transmission for not disrupting the data transmission from node 1 since it is within node 2's carrier-sense range. When node 1 transmits, only node 5 can transmit concurrently. The chain therefore has a maximum expected utilization of $\frac{1}{4}$ of each node's link capacity. However, simulations done by Li et al. [LBC⁺01] only found a utilization of $\frac{1}{7}$ because the IEEE 802.11 MAC was unable to discover the optimal transmission rate on its own. In a more complex set-up with a lattice topology, the same authors found a channel utilization of $\frac{1}{12}$. One effect of the lower throughput for multi-hop paths is the unfairness involved between flows with different path lengths; comparable flows will not have the same predictable performance end-to-end. Fairness and predictability are discussed in more detail in Section 3.

Another aspect affecting the capacity, especially in multi-hop networks, is the routing. Different routing protocols have different approaches to the routing; generally *proactive* versus *reactive* routing (as discussed in Section 1.2). Perkins et al. [PRDM01] compared the performance of DSR and AODV, which are both reactive routing protocols, and found differences for different load levels and mobility.

Most research on capacity either considers a static ad hoc network, or identical nodes with a uniform mobility model. The assumptions are, however, relaxed by a recent analysis on capacity by Wang et al. [WSGLA08]. Wang et al. present a framework for the computation of throughput capacity in random wireless ad hoc networks. A different scenario appear when ad hoc nodes may communicate with wireless infrastructure nodes that are connected by high capacity links. The scenario is very relevant today, having several WLAN hot-spots around, and more WMNs deployed. This problem was addressed by Zemlianov and de Veciana [ZdV05].

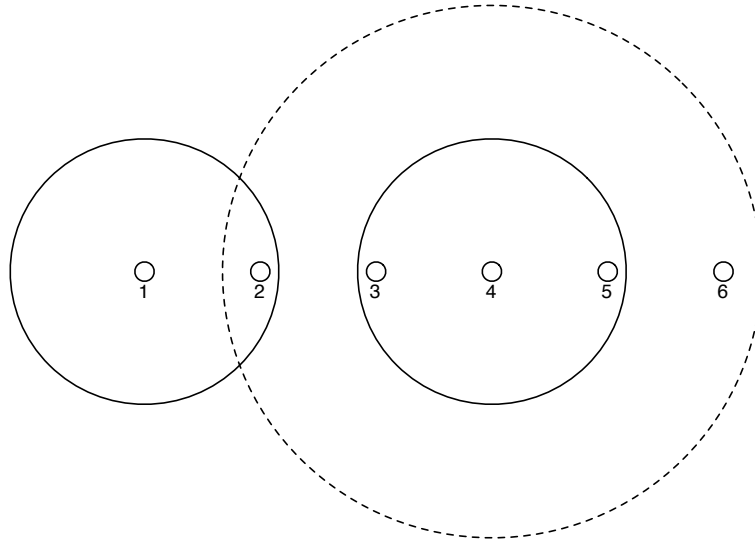


Figure 4. Interference in a chain of nodes [LBC⁺01]

1.3.1 Discussion

Since standard parameter values are used for the transmission and carrier-sense ranges, the most relevant problem for this thesis is how the throughput depends on the number of hops made; as shown by Li et al. [LBC⁺01], multi-hop flows will experience a reduction in the effective capacity. An effective capacity that depends on the source and receiver locations gives less predictability for the users. A MANET implicitly imply an unpredictable performance due to the dynamic environment. But predictability should be targeted when the number of hops are known to give users more equal performance for the same services. The importance of predictability is emphasized in Section 3, and targeted by Paper D in Part II of this thesis.

1.4 Wireless Mesh Networks

A typical WMN is a wireless communication network that provides Internet access for mobile nodes. The WMN consists of two types of nodes; mesh routers and mesh clients. The mesh routers have minimal mobility (most often stationary) and can be more heavy-weight in terms of routing capability. Their objective is to create a wireless communication backbone for the mesh clients, which can be much simpler with regards to hardware and software implementations.

WMNs have gained a lot of interest lately based on their many application scenarios. Low deployment cost, robustness, and easy network maintenance make WMNs beneficial in home, enterprise, and metropolitan networking [AW05]. Figure 5 depicts a typical WMN scenario where a wireless mesh backbone provides wireless coverage and Internet access for mesh clients. The underlying wireless network technology can

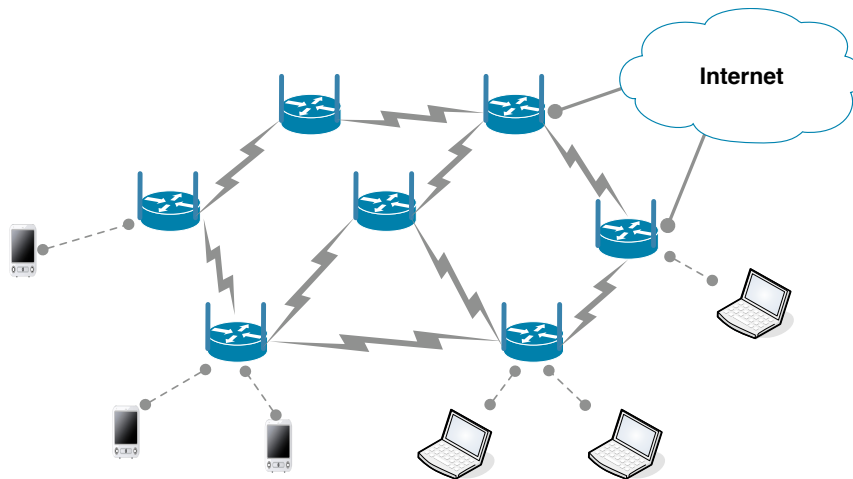


Figure 5. An infrastructure WMN [AW05]

be different for different WMNs. Existing technologies like IEEE 802.11 and IEEE 802.16 [IEE04], known as Wi-Fi and WiMAX, can be used, and their standardization groups are actively working on specifications for WMNs.

WMNs can be classified into three types based on how mesh routers and mesh clients interact: infrastructure, client, and hybrid WMNs. In infrastructure WMNs, the WMN operates as a backbone for clients. Figure 5 is an example of an infrastructure WMN. Clients can either connect directly to the mesh backbone, or they can connect through gateways using a different radio technology (i.e. a cellular mobile phone through a base station). Client WMNs are the same as conventional ad hoc networks. Hybrid WMNs are a combination of infrastructure and client mesh networks. This type of WMN is present when for example a MANET uses a wireless infrastructure network for Internet access. The state-of-the-art within WMNs are provided in a survey by Akyildiz and Wang [AW05].

One scenario in the context of this thesis is where MANETs use a WMN to connect to the fixed network. Then the WMN must also take part in the integration between MANETs and fixed networks. From a QoS perspective, it would add another challenge to the end-to-end support. With a DiffServ approach the WMN could be seen as a different domain to cross for communication between the MANET and the fixed network.

2. Quality of Service in IP Networks

Traditionally, IP networks have offered a best-effort service for data transmissions. However, as more demanding application types like Voice over IP (VoIP) have evolved, better guarantees are needed. QoS is defined by the Telecommunication Standardization Sector of the International Telecommunication Union (ITU-T) as the collective

effect of service performance which determine the degree of satisfaction of a user of the service [ITU94].

It is especially important to provide QoS in capacity constrained networks like wireless networks. When over-provisioning of capacity is not possible, the available capacity must be controlled for an improved utilization.

The QoS can be provided at different degrees. Hard guarantees come from end-to-end per-flow reservations, where a portion of the available resources are dedicated for a restricted number of flows. Softer guarantees are given in differentiation schemes, where a flow's performance must be seen in relation to other flows and is often stated within performance bounds. Flows are then a part of an aggregate which is given differentiated treatment from other aggregates (i.e. traffic classes or service classes). QoS can be assessed from subjective or objective measures; qualitative testing is done through subjective user testing, while quantitative measures are metrics like bandwidth, loss, delay and jitter. For this thesis, only the quantitative parameters are considered.

QoS is in a broader sense a part of TE. TE considers how paths should be selected to optimize resource utilization and traffic performance in the network, which include the aspect of resource allocation. Different QoS models have different means to allocate resources. In this section, such models are presented and discussed. Two QoS architectures designed for fixed networks are Integrated Services (IntServ) [BCS94] and Differentiated Services (DiffServ) [BBC⁺98]. These architectures approach the QoS support differently and have been used as basis for other proposals, including proposals for MANETs. IntServ and DiffServ are presented in Section 2.1.

Important mechanisms for QoS provisioning are admission control and congestion control. The admission control restricts the amount of traffic in the network. Only flows with requirements that do not degrade the service of the existing traffic are admitted into the network. However, traffic can be bursty, and the topology may change quite rapidly due to mobility, necessitating congestion control. One approach is to make the source adapt to the new conditions. But not all sources may adapt due to their application requirements, and it might also be difficult to find new acceptable levels. Another approach is to preempt flows. In wireless conditions, preemption might be the only solution to avoid a degraded service for many of the existing flows. Admission control and congestion control mechanisms are presented in Sections 2.2 and 2.3, respectively. Then QoS solutions for MANETs are considered in Section 2.4. Here, QoS in both the MAC and network layers are considered. Section 2.5 presents some QoS proposals interconnecting MANETs and fixed networks. In Section 2.6, this thesis is discussed in relation to the QoS solutions presented.

2.1 QoS Architectures for Fixed Networks

In this section, two QoS architectures, which are of most relevance to the work done on QoS in MANETs, are presented: IntServ and DiffServ. The IntServ architecture provides hard guarantees with end-to-end per-flow reservations, while the DiffServ architecture improves the scalability and offers softer guarantees by differentiating among aggregates of flows. Both of these models were defined for fixed networks, but

elements from these have been used in architectures for wireless networks as well. Section 2.4 considers QoS in MANETs, in where ideas have been taken from both the IntServ and DiffServ architectures.

Integrated Services

In 1994, the IETF proposed the QoS architecture IntServ to guarantee QoS in a communication network. IntServ provides per-flow guarantees end-to-end by reserving resources along the path. IntServ uses the Resource Reservation Protocol (RSVP) [BZB⁺97] to reserve resources in each intermediate node, which requires an admission control for each node. In the RSVP reservation message, the flow's traffic and reservation specifications are included. The Traffic Specification (Tspec) describes the requesting flow's traffic by token bucket parameters. To get the requested service level, the flow must conform to the given Tspec.

IntServ provides two services in addition to best-effort: Controlled-Load [Wro97] and Guaranteed QoS [SPG97]. The Controlled-Load service gives a flow the same service as best-effort under *unloaded conditions*. Exceeding the agreed Tspec may cause degraded service. The Guaranteed QoS service assures a level of bandwidth, and gives a delay-bounded service with no queueing loss if the flow conforms to the Tspec. Non-conforming packets are treated as best-effort packets. Without any activity, a flow's reservations are released for others to use.

The main disadvantage with IntServ is the scalability; each node must maintain a high number of reservations in networks like the Internet. One solution to this problem is to aggregate multiple RSVP reservations (Aggregated RSVP (aRSVP)) to one reservation in the core network [BIFD01].

Differentiated Services

The per-flow reservations in IntServ do not scale to large networks. DiffServ increases the scalability by providing an architecture to allow per-class classification and differentiated treatment. Within this architecture, each network operator can define their own policy on how different traffic types are aggregated and treated. At the arrival to a *DiffServ domain* (i.e. group of nodes with common defined DiffServ policies), each packet is classified and given a Per-Hop Behavior (PHB), which defines the forwarding behavior through the domain. Figure 6 shows the communication from one DiffServ domain, crossing a second, and reaching a third DiffServ domain. Classification at the domain boundary, by the boundary node (BN), simplifies the complexity in the core nodes (CNs). The domain's routers differentiate among the PHBs by buffer management and packet scheduling mechanisms. Optionally, the packets may be conditioned for compliance with defined traffic profiles; there might exist Service Level Agreements (SLAs) with traffic profile specifications when communicating to, and between, different DiffServ domains. At least one mechanism should exist to control the load to not degrade the network's services.

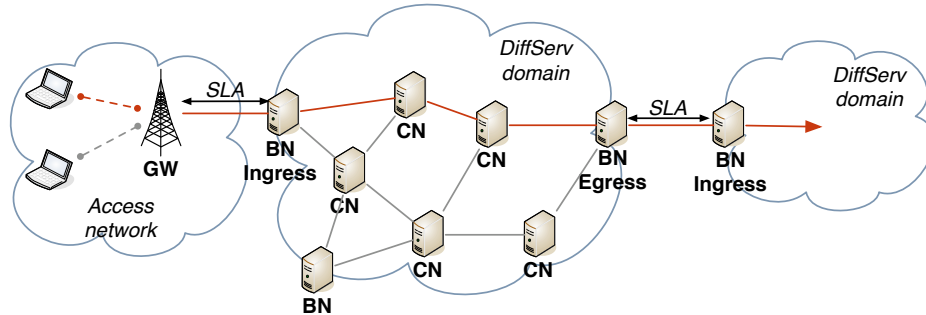


Figure 6. Communication through a DiffServ enabled domain

All packets having the same PHB belong to a Behavior Aggregate (BA). The PHB is encoded into the IP header's 6-bit long Differentiated Services Code Point (DSCP). Among the possible 2^6 PHBs, there is a set of defined PHBs: Expedited Forwarding (EF) [DCB⁺02], Assured Forwarding (AF) [HBWW99], and the default PHB (i.e. best-effort). The EF PHB intends to provide low delay, jitter and loss, and is suited for strict real-time applications. The AF PHB group defines four independent AF classes, in where three drop precedence levels can be chosen, giving a total of twelve AF classes. In case of conflicts, a packet's importance is decided by the AF class and the drop precedence within the class.

One disadvantage with the DiffServ approach is that it is difficult to predict the end-to-end performance when traversing domains with different policies. Also, DiffServ has been criticised on the cost it adds on aspects like management, configuration and service level specifications between different parties [Dav03]. This is especially true for networks with much excess capacity. However, like Davie argues in [Dav03], the ability to differentiate traffic with different characteristics, and the growth of more capacity constrained environments like wireless networks, which increases the focus on utilizing the existing capacity, may increase the operators' incentive for more DiffServ deployment in the future.

2.2 Admission Control

The admission control checks whether a network can accommodate a new data flow, based on the network's traffic and the flow's QoS requirements, without affecting already admitted flows. Typical QoS requirements are metrics like loss, delay, and bandwidth. Admission control is a part of the overall goal for TE; utilize the network resources and improve the traffic performance.

There are different means to restrict traffic from entering a network. Because of the scalability issues, most approaches have been on distributed admission control; in MANETs, which are distributed by nature, this is especially important. The admission

control must consider both local and non-local conditions. Taking only the local conditions into account may lead to a deterioration of other nodes' performances.

Admission control schemes can be grouped by where the admission control takes place; generally the schemes have either an admission control at each hop, or an admission control only at the end node. The admission control proposed in this thesis uses a source-based approach. In MANETs, the admission process has typically involved end-to-end signaling, either through the routing protocol or by dedicated signaling messages (i.e. probes). The End-to-End Measurement-Based Admission Control (EMBAC) schemes send probes along the path, and compare the received metrics like available bandwidth, loss and delay to admission control thresholds at the end node [Mas04, KKZ00, EKR00]. Different proposals differ in amount of state information maintained at each node. Breslau et al. [BKS⁺00] gives a study of architectural issues for endpoint admission control schemes. However, the schemes were not specifically considered in the context of wireless networks.

Finding loss and delay is done directly on the probes, while the available bandwidth is more complex to find. Measuring the interference level provides an estimate on the available channel capacity, but the problem is to estimate how the new flow will affect the neighboring nodes. Yang and Kravets [YK05] define a neighbor-carrier-sensing threshold, which is lower than the ordinary carrier-sensing threshold. This threshold makes them able to estimate activity within their neighbors' carrier-sense ranges as well.

A more holistic view was presented by the Traffic Engineering for Quality of Service in the Internet, at Large Scale (TEQUILA) project for fixed networks [Inf]. The TEQUILA approach [MCG⁺03] aimed at efficiently resolve the trade-off between utilization of the network resources and QoS deterioration in DiffServ IP networks. The approach is strongly policy-driven, and deploys an admission control at both the subscription and invocation level through feedback of anticipated demand and measures of actual status, respectively. Implementing it in MANETs is more complex, but incorporating more policy-driven admission decisions in MANETs are of importance.

2.3 Congestion Control

In large-scale communication networks like the Internet, it is important to have some sort of congestion control for network stability, throughput efficiency, a predictable low delay, and fair resource allocation to network users [MHT07].

In MANETs, mechanisms are needed due to topology changes and consequent changes in traffic pattern. One approach is to interpret network information like loss and delay, while another approach is to send explicit status messages in the network. Implicit messages, like lost ACKs, may indicate a congestion somewhere along the data path. The interpretation of the lost packets can, depending on the application and transport protocol, lead to a decreased data transmission rate from the source. TCP is an example of this approach. Explicit messages reports on congestion, or situations soon to become a congestion. Here, ECN is an example. Since both TCP and ECN are methods to handle congestion, and are used in this thesis, they are described in more detail below in this section. In addition, the Datagram Congestion Control Protocol

(DCCP) is presented as an alternative to TCP.

Transmission Control Protocol

TCP provides reliable transmission of data by retransmitting lost packets and an ordered delivery to the upper layer, along with a window-based congestion control [CDS74, Inf81]. TCP consists of four algorithms, *slow start*, *congestion avoidance*, *fast retransmit*, and *fast recovery* [APS99], which were all developed by Jacobson [Jac88, Jac90]. The number of transmissions is guarded by a sending window, which is the minimum of the receiver-side limit on the amount of outstanding data and the Congestion Window (CWND). The slow start increases the CWND exponentially until the Slow Start Threshold (*ssth*), from where the congestion avoidance takes over and increases the CWND linearly through the Additive Increase Multiplicative Decrease (AIMD) algorithm. Duplicated ACKs indicate dropped packets and congestion. Upon congestion, the CWND is halved before doing a fast retransmit of the lost segment(s) and going into congestion avoidance again, which is the fast recovery algorithm (the Reno version [APS99]). Figure 7 provides an illustration on how the CWND can vary with time. Slow start ends when the $ssth_1$ is reached. When a packet loss occurs at time 2, the *ssth* is set to half of the current window (here: set the *ssth* to $\frac{C_1}{2}$) before doing fast retransmit and fast recovery. The CWND then continues in congestion avoidance. Under fast recovery, an ACK on the lost segment(s) is required within the timeout to avoid a new slow start. Many TCP extensions have been proposed, and [DBEB06] gives a roadmap for all the Request For Comments (RFCs) related to TCP.

Datagram Congestion Control Protocol

The reliable transport services with in-order delivery and congestion control, which TCP provides, are not suitable for delay sensitive application types due to arbitrarily long delays. The basic transport layer protocol for such real-time applications is the User Datagram Protocol (UDP) [Pos80], but UDP does not include any congestion control. DCCP [KHF06] was designed by IETF to provide congestion control, including ECN support, for time-constrained applications without TCP's arbitrarily long delays. Currently, there are two defined congestion control mechanisms (Congestion Control Identifiers (CCIDs)): TCP-like congestion control [FK06] and TCP-Friendly Rate Control (TFRC) [FKP06]. Streaming media applications that require a relatively smooth sending rate is best supported by TFRC.

Explicit Congestion Notification and Active Queue Management

The ECN mechanism gives explicit messages about network congestion to the end nodes. The IP-header's ECN bits are used for the notification. The sources receiving notifications must then behave accordingly by e.g. adapt its transmission rate or preempt its session.

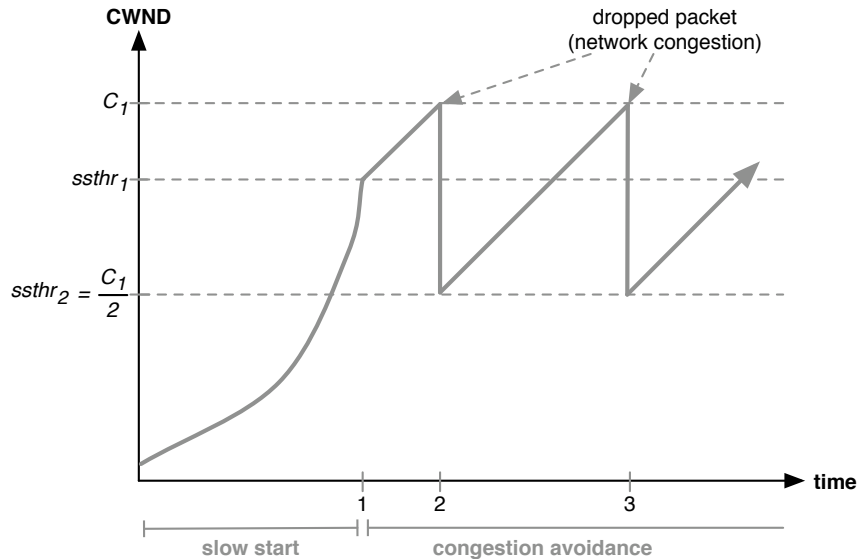


Figure 7. Variation of the CWND in the TCP congestion control

To discover congestion in the network, ECN depends on Active Queue Management (AQM) to supervise and discover congestion before they appear. One example of such a strategy is the binary feedback scheme DECBIT for congestion avoidance [RJ90]. When the calculated average queue length exceeds a threshold, a congestion-notification bit in the packet's header is set. Another method for detecting congestion at the gateway is provided through the Random Early Detection (RED) mechanism [FJ93]. RED computes a weighted average between current and average queue length. RED can be used together with both cooperative and non-cooperative transport layer protocols (e.g. TCP and UDP, respectively). Packets are marked or dropped at random when the weighted average is above the threshold. A flow's marking probability will thus be proportional to the flow's share of the gateway's resources.

In Figure 8, the source is notified about a congestion on its path to the destination. The marking in the figure follows the IETF RFC on incorporation of ECN to TCP and IP [RFB01]. The cooperation between TCP and ECN is considered by Floyd in [Flo94], while [RFB01] considers ECN in more general IP networks. In 2004, there were not many ECN capable servers deployed in the Internet [MAF05]. Based on such information, and to give incentives for more ECN deployment, Kuzmanovic discusses the power of ECN in [Kuz05].

2.4 QoS Architectures for MANETs

QoS in MANETs has been approached from different angles in the research. In this section, different QoS architectures are presented. In addition, a QoS aware MAC protocol and QoS routing are also presented.

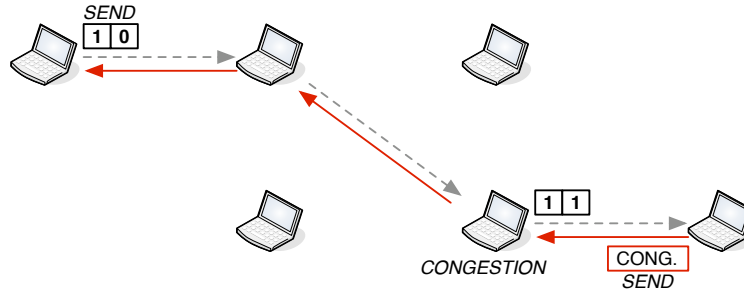


Figure 8. Using ECN to notify on a congestion

One QoS aware MAC protocol is the IEEE 802.11e standard [IEE07], which is an extension of the IEEE 802.11 standard. IEEE 802.11e defines a new coordination function, the Hybrid Coordination Function (HCF), which defines two new medium access mechanisms: the HCF Controlled Channel Access (HCCA) and the Enhanced Distributed Channel Access (EDCA). In the context of MANETs, the contention-based EDCA is of most relevance. But, this layer 2 solution requires mechanisms on higher layers for end-to-end QoS support. However, since it can be an improvement to the architecture presented by this thesis, the EDCA is presented in more detail below.

QoS routing gives end-to-end QoS support by requiring more from a route than only connectivity. QoS routing is either an extension to an already existing routing protocol for MANETs, or a new routing protocol built for the purpose of providing QoS along the path. An example of both types are presented below.

Most QoS architectures for MANETs have been based on either resource-reservations like the IntServ architecture, or differentiated behaviour like the DiffServ architecture. In this section, both IntServ and DiffServ based schemes are reviewed. INSIGNIA (an IP-based Quality of Service Framework for Mobile Ad Hoc Networks) [LAZC00] is based on ideas from IntServ, and provides in-band signaling for bandwidth reservations. This thesis, on the other hand, belongs to the DiffServ-class of QoS architectures. Examples of schemes that provide differentiated treatment are Service Differentiation in Stateless Wireless Ad Hoc Networks (SWAN) [ACVS02], Flexible Quality of Service Model for Mobile Ad Hoc Networks (FQMM) [XSLC00], and Multipath QoS Routing Protocol of supporting DiffServ (MQRD) [LC05], along with schemes proposed by Haq et al. [HMBT04] and Xu et al. [XTB⁺03].

MQRD builds on a routing protocol that is a combination of AODV and DSR for providing node-disjoint multi-paths. It is further integrated with DiffServ to provide QoS along each multipath. Haq et al. incorporate the DiffServ approach by separating the edge and core functionality. Each node in their model implements both the edge and core functionality for traffic conditioning and packet forwarding according to the PHB, respectively. Xu et al. define four phases: adaptive bandwidth management, QoS routing, Call Admission Control, and congestion control. Admission is based

Table 1. User Priority to Access Category mapping [IEEE07]

UP	Designation (informative)	AC
1	Background	AC_BK
2	Background	AC_BK
0	Best-effort	AC_BE
3	Best-effort	AC_BE
4	Video	AC_VI
5	Video	AC_VI
6	Voice	AC_VO
7	Voice	AC_VO

on bandwidth estimations and routing information from the LANMAR [GGH00] protocol. Congestion is discovered in the bandwidth estimation process.

INSIGNIA, SWAN, and FQMM are presented in more detail later in this section.

QoS aware MAC: The IEEE 802.11e EDCA

The IEEE 802.11 DCF treats all packets equally, providing a best-effort channel access to the packets. To provide QoS on the MAC layer, the MAC protocol should distinguish high priority packets and give them differentiated treatment from lower prioritized packets. The EDCA gives differentiated treatment to four Access Categories (ACs) to serve eight User Priorities (UPs). These UPs are mapped to the ACs according to Table 1, where the UPs are listed from lowest to highest priority. There are three time values that differentiate the ACs:

- **Time after transmission**

The Arbitrarily IFS (AIFS) defines the time an AC has to wait after the channel is sensed to be free. DCF defined a DIFS independent of traffic type and priority. The EDCA defines an AIFS for each class, where a lower value gives less time before backoff or transmission.

- **Time in backoff**

The Collision Window (CW) defines the time the sender stays in backoff. In the DCF, CW_{min} and CW_{max} were both independent of traffic type. The CW_{min} and CW_{max} are in EDCA dependent on the AC. A lower CW will thus give a shorter backoff time before a possible transmission.

- **Time in transmission**

EDCA offers an opportunity to transmit multiple frames consecutively without contention from other stations, i.e. Transmission Opportunities (TXOPs). The TXOPs differ based on the AC, and are given by the $TXOPLimit$ parameter specified for each AC. A value equal to zero allows only one frame, while a higher value may indicate multiple consecutive frames. The TXOP can increase the network throughput, especially with an Access Point (AP) present; the AP can then assign TXOPs dynamically, instead of using the predefined value.

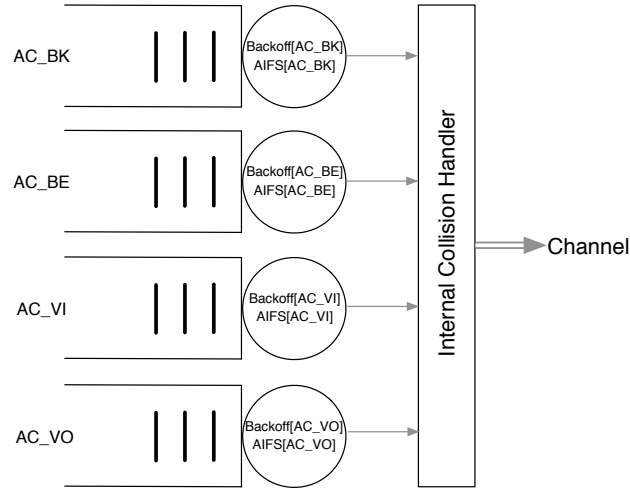


Figure 9. IEEE 802.11e transmission queues

Figure 9 illustrates the EDCA, where each AC queue roughly deploys a DCF with different AIFSs and backoff times. The internal collision handler resolves collisions between the different ACs at the same node. Figure 2 showed the basic operation of CSMA/CA in DCF. If EDCA was deployed, and node C was given priority over node B by for example a shorter AIFS, the probability of having node C transmitting first would increase.

There has been an increase in research on the IEEE 802.11e standard after the release. One example is a performance study by Calafate et al. [CMM04]. They found that the high priority traffic maintained a steady throughput independently of the lower priority traffic load for a static scenario, and kept most of its effectiveness in mobile scenarios as well.

QoS only on the MAC layer is not enough to fully support QoS in the network. IEEE 802.11e networks will also need network control mechanisms to avoid network overload. Therefore, the IEEE 802.11e will only be one part of a total QoS architecture for supporting the requesting services. IEEE 802.11e networks could interconnect with DiffServ enabled networks as well by mapping the ACs in EDCA to PHBs in DiffServ. One initial proposal is presented by Elmangosh et al. [EAS07].

QoS Routing

QoS routing is to find a path from source to receiver that can support the QoS requirement(s) defined by the session. Typical QoS metrics used in the path discovery are bandwidth and delay.

Other metrics like the Signal-to-Noise Ratio (SNR) and power level, which indicate the link quality and network stability, respectively, are also used. Hanzo-II and Tafazolli [HIT07] give a comprehensive presentation of different QoS routing protocols.

The survey also presents metrics used, problems, and design consideration in the QoS routing process. QoS routing is either based on an extension to an already existing routing protocol, or is a new protocol built for the purpose of providing QoS along the path. An example of both types is presented here: the QoS extension to AODV by Perkins and Belding-Royer [PBR01], and the Core-Extraction Distributed Ad Hoc Routing (CEDAR) algorithm by Sivakumar et al. [SSB99].

QoS in AODV is provided by QoS extensions to the path setup messages, along with a routing table extension. The included QoS extensions specify the requirements intermediate nodes must be able to meet in order to forward RREQ or RREP. The available QoS parameters are bandwidth, delay and jitter, along with a traffic class definition related to DSCP. See the description of the original AODV in Section 1.2.2 for the path setup. To keep the extra state information, the routing table is extended with the new information. If the requested QoS can no longer be provided along the path, the node detecting this must send a QoS-lost message to the source of the requested QoS. This QoS extension by Perkins and Belding-Royer is easy to deploy with little overhead in AODV based networks. But the performance will depend on scenario and available bandwidth estimation.

Another approach is CEDAR, which creates and maintains a network of core nodes chosen for dissemination of routing and bandwidth information. Each node selects the neighbor with the highest degree in its one-hop neighborhood as its core node to be associated with (i.e. its *dominator*). By maintaining a set of core nodes, there is less route computations and state management in the network. Also, unreliable broadcasts with hidden and exposed nodes are avoided through RTS/CTS in unicast between the core nodes. A standard routing protocol like AODV can be used within the core network. Each core node knows its local topology and relatively stable links to other core nodes. Bandwidth information is propagated through the core network by increase and decrease messages (i.e. *waves*), where decrease waves move faster. QoS requests from a non-core node are sent to its dominator, which handles the further propagation. CEDAR can increase the scalability with the creation of a core network, and link status messages are propagated fast for up-to-date information. However, finding the minimum dominating set is hard, so approximations are used. The available bandwidth estimation is assumed to be provided by the MAC layer.

INSIGNIA

INSIGNIA [LAZC00] is an IP-based QoS framework for MANETs. It provides soft-state reservation of bandwidth specified by the applications. The application specifies the requested bandwidth (minimum (*minBW*) and maximum (*maxBW*) bandwidth) and sends the request towards the destination. At each intermediate node, admission control is performed to see if the requested level can be reserved. If not, *maxBW* is degraded to *minBW* or best-effort, while *minBW* is degraded to best-effort. Then, the destination reports about what QoS level the network can provide on the particular path.

The reserved resources are reserved in soft-state, which means that refresh messages are needed to keep the reservations alive. The reservations' timeout duration will thus indicate a trade-off between signaling overhead and adaptability in the network. A lower timeout duration means more overhead in the network, but better adaptability. Topology changes may need restoration by the underlying routing protocol, which can lead to new reservation signaling. If the requested QoS can not be provided on the new path, the source application must adapt if possible.

INSIGNIA is a protocol with low complexity, which is independent of the routing and MAC protocol, but the end-to-end QoS performance will be dependent on the MAC implemented. The decoupling from the routing protocol allows INSIGNIA to use different routing schemes in different contexts. The performance was evaluated with the routing protocols AODV, DSR and Temporally Ordered Routing Algorithm (TORA) by Lee et al. [LAC01]. The signaling in INSIGNIA is performed in-band, i.e. the refresh messages are piggybacked onto the data packets. This lowers the overhead, and gives fast responses on the network dynamics.

The reservations are, as mentioned, soft-state, which avoids session time reservations. However, the resources can be held unnecessarily long during the reservation stage since resources are reserved along the path without knowing if the entire path can support the requested bandwidth. This may block other requests, which have available bandwidth on their entire path, thus wasting the resources for the duration of the timeout.

To avoid this resource waste, Xue et al. [XSA03] proposed Adaptive Reservation and Pre-allocation Protocol (ASAP). In ASAP, others can use the reserved resources (*soft reserved*) until the destination confirms that the entire path can do the same (*hard reserved*). Despite of the per-flow QoS performance, aspects such as admission control based on bandwidth estimations, along with resource reservations, make INSIGNIA less attractive in MANETs.

SWAN

SWAN [ACVS02] is a stateless QoS architecture, which differentiates between real-time traffic and best-effort traffic. For the real-time class, the source probes the network before admission. At each intermediate node, the available bandwidth is checked, and the probe's bottleneck bandwidth information is updated if less is available. No bandwidth is reserved along the path. The best-effort traffic is controlled by the AIMD algorithm, which is based on the delay at the MAC layer. Mobility, causing contentions, and false admissions (e.g. when multiple sources probe the network simultaneously to find enough bandwidth available) are solved with ECN. Two approaches are proposed for situations with congestion; mark all packets or only a subset based on the sessions' durations. Figure 10 shows the SWAN model, where the admission controller handles the probing, and the rate controller shapes the best-effort traffic according to the MAC delay.

The most appealing aspect of SWAN is its stateless and source-based admission control approach, which lower the processing in intermediate nodes. Admission

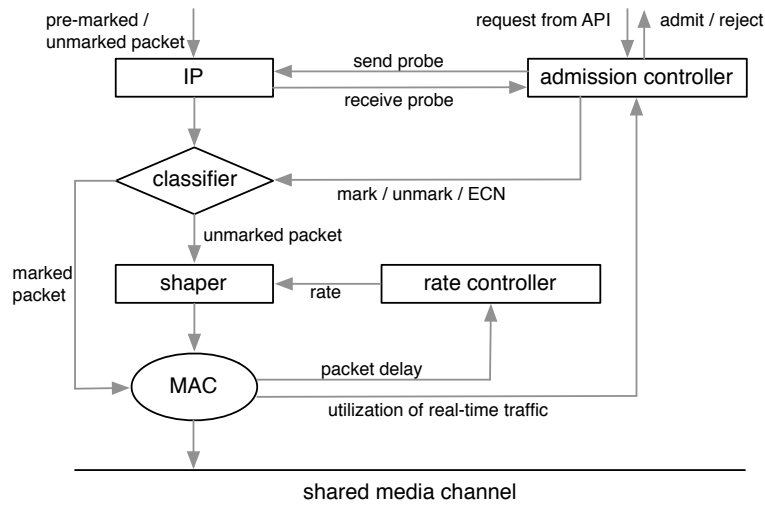


Figure 10. The SWAN model [ACVS02]

control, along with ECN and rate control for the low priority traffic, ensure protection for the high priority traffic.

One problem is to set the correct initial rate for the best-effort traffic. The initial rate is decisive for the performance, but no procedure is suggested in the paper. Another problem is how to measure the local resource availability. Due to the carrier-sense range, it is difficult to find good estimates of the real-time traffic already present. SWAN takes a conservative approach when defining the upper threshold of admission, but will still be dependent on the estimates. Also, support of only two classes is not enough for serving a wide spectrum of application needs. In an extension to SWAN, Crisóstomo et al. [CSNV05] relax the number of classes to four. The highest prioritized class will use probing, and the lowest prioritized class is controlled by AIMD, as in SWAN. For the two classes in between, both probing and AIMD control are employed.

Despite some drawbacks, the SWAN model is an attractive approach to QoS in MANETs. Especially with the interconnection possibilities with DiffServ enabled fixed networks, which is described in more detail in Section 2.5. Crisóstomo et al. also extend the SWAN model with multipath support. This is realized by a routing protocol extension of AODV, called AOMDV [MD01]. Sridhar and Jacob [SJ04] remove the probing phase by including bandwidth information in the routing protocol (i.e. QoS routing). The drawback is that the underlying routing protocol must then be altered to carry bandwidth information.

A solution in close relations to SWAN is the Cross-layer Architecture for DiffServ (CLAD) in MANETs by Zhou et al. [ZML05]. For the admission of real-time traffic, CLAD probes the network for available path capacity, and uses an adaptive rate controller for the best-effort traffic. A QoS enabled routing scheme that tries to find

the least busy route is used. Also, four service classes are defined and differentiated by the scheduler. However, it is hard to tell to which extent the service classes are differentiated since only two classes are tested in the simulations. But what is mainly new compared to SWAN, is a method for setting the initial rate of the best-effort traffic, which depends on the predictable saturation level of the channel. The rate controller will then more quickly adapt to the link state.

FQMM

FQMM [XSLC00] combines the IntServ and DiffServ architectures to get the best from both approaches; the per-flow granularity in IntServ is combined with the per-class granularity in DiffServ. IntServ provides good QoS support for individual flows, but due to the scalability problems, only the highest prioritized flows are controlled through IntServ, while the rest use DiffServ. To avoid any scalability problems, it has to be assumed that the highest prioritized flows constitute only a small fraction of the traffic. FQMM can use any routing protocol that can find the available bandwidth along a path.

FQMM will initially inherit the scalability properties from DiffServ, but a problem appears when the assumption of little high priority traffic does not hold. Then, IntServ's scalability problems become more evident on the network as a whole. In order not to overload the network, traffic is conditioned at ingress nodes (nodes will have dynamic roles) according to the traffic profile. The aim is to keep the differentiation between classes predictable and consistent, while also adaptive to the dynamics of the network. However, implementing both IntServ and DiffServ in a MANET makes FQMM look less appealing due to the complexity involved. Also, only a preliminary simulation study is provided by the authors, so more analysis are needed.

2.5 QoS over Interconnected MANETs and Fixed Networks

Providing QoS in MANETs is important for the internal communication. However, a wireless node will often need communication with external fixed networks, for example to access the Internet. To support QoS end-to-end in such scenarios, both domains must provide QoS and understand each other. Figure 11 shows a typical interconnection scenario between a MANET and a fixed network where the communication goes through a gateway (GW). Most of the research on QoS in MANETs has only covered the communication within the MANET, but some proposals have considered QoS schemes for interconnected MANETs and fixed networks. PYLON [MK03] is an architectural framework for QoS interconnectivity, while a group of other proposals have extended the SWAN architecture. Using SWAN for the interconnectivity inherits the benefits and imitations of SWAN described earlier. In [CSNV05] and [DR08], SWAN is extended to operate along an access network for Internet access. These approaches, along with PYLON, are reviewed in this section. Other proposals that extend SWAN networks for interconnectivity are [DR04b] and [SK05].

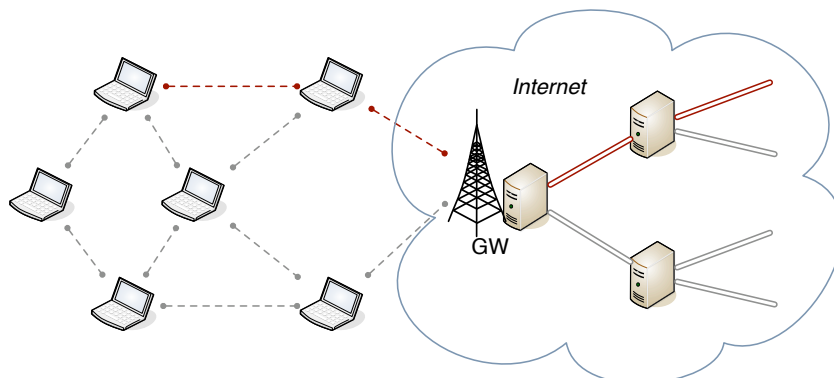


Figure 11. Interconnecting a MANET with a fixed network

PYLON

PYLON [MK03] is an interconnection framework that opens for several QoS architectures in the ad hoc network, along with a fixed network infrastructure of either DiffServ or IntServ. The main candidates for providing QoS in the ad hoc network are INSIGNIA, FQMM, and SWAN, according to the authors.

A GW will in PYLON broadcast its presence and offer its services to ad hoc networks. The GW keeps traffic related information about earlier ad hoc networks, and can based on this information suggest a customized SLA and Traffic Conditioning Agreement (TCA) to a new ad hoc network. Leaving most configurations to the access network saves processing and memory in the ad hoc network. The ad hoc network must keep a detailed list of DSCP codes for support of DiffServ traffic, and to ease the mapping in the GW, which contains the access network's "in-service DSCPs". PYLON supports RSVP, and uses aRSVP to hide microflow information and increase the scalability. One problem with aRSVP is how to determine the overall size of the reservations and how to update the reservations, which becomes a trade-off between simplicity and scalability. The problems concerning RSVP are, however, removed if a complete class-based QoS architecture is used in the ad hoc network, for example the candidate SWAN. The authors only present a high level view of PYLON, so more analysis are needed to show the benefits of such an approach.

Extensions to SWAN

Crisóstomo et al. [CSNV05] do not only add, as mentioned before, support for more differentiated classes and multipaths to SWAN, but they also add interconnection support for SWAN networks. For a request to the fixed network, the GW cooperates with a QoS broker for the bottleneck bandwidth (BB) information in the fixed network. The idea is to use the same DSCP codes in both the ad hoc network and the infras-

structure network to facilitate the mapping in the GW. To simplify the view from the infrastructure side, due to the unstable wireless links, the path between the GW and a node in the ad hoc network is considered as a virtual link with capacity equal to the BB along the same path. The GW will also monitor the flows to discover any ECN marked packets for a quick response back to the source.

Domingo and Remondo propose a modified SWAN scheme in cooperation with a DiffServ enabled fixed network (Differentiated Services - SWAN (DS-SWAN)) initially in [DR04a], and later improved in [DR08]. The ingress node in the DiffServ domain monitors the loss and sends a QoS-LOST message back to the ad hoc network if the loss is too high. A QoS-LOST message can also be initiated from the destination, e.g. in the case when application specific thresholds are surpassed.

When a node in the ad hoc network receives a QoS-LOST message, it reduces its AIMD parameters; the increment rate is decreased, the decrement rate is increased, and the minimum leaky bucket rate for best-effort traffic is decreased. However, all these parameters have minimum and maximum values to avoid starvation of the best-effort traffic. The inverse happens if no QoS-LOST message is received during the last T seconds. The QoS-LOST message will extend the ECN mechanism in the original SWAN; the dynamic AIMD parameters will lower the delay and loss experienced for flows requiring QoS. Placing the monitoring in the fixed part of the network, as PYLON, will avoid unnecessary battery consumption in the ad hoc nodes. To further enhance the QoS, a modified AODV routing protocol (Service Differentiation - AODV (SD-AODV)) is proposed to avoid congested nodes in the route request phase.

2.6 Discussion

The work in this thesis has focused on developing a QoS architecture for MANETs. While developing a solution for the MANET, it is important to have the interconnection with fixed networks in mind since Internet access is important in today's networking. This was the motivation for extending a QoS architecture designed for fixed networks into the MANET domain.

Two viable alternatives to use were IntServ and DiffServ, each with their own advantages and disadvantages. Resource reservations need available bandwidth estimations and per-flow state information in all nodes along the path. It provides good QoS, but is less attractive in the context of MANETs. Available bandwidth estimations have inaccuracies, and the non-scalability of storing state information should be avoided in MANETs. The approach is exemplified by INSIGNIA in MANETs. One solution might be to allow only a small fraction of the traffic to use the IntServ mechanisms as in FQMM. Another is to aggregate RSVP flows for better scalability as in PYLON.

DiffServ approaches QoS by classifying flows into a set of behaviors, or classes, and gives differentiated treatment among these classes. Differentiating on a set of classes, avoiding per-flow awareness, is less complex in the constrained and varying wireless networks. In addition, the complexity is moved towards the edge, making the forwarding less comprehensive for often low powered intermediate nodes. Such issues give more incentives for using DiffServ in MANETs. Consequently, using DiffServ should provide differentiated treatment to the service classes in the MANET,

and give a more seamless operation with fixed networks. Another issue is the more policy-driven approach of DiffServ, which is of importance when several different domains are connected for interdomain communication.

The shadow class concept proposed in this thesis is a source-based admission control approach without any challenging bandwidth estimations. SWAN, which is the most comparable QoS architecture for MANETs, uses probing to check the conditions along the path. In the shadow classes the real data flow is transmitted to better see all effects of the new flow. Congestion is solved by ECN, which is also used in SWAN. For a number of marked packets, the flow becomes preempted since a possible degradation in prioritization would still impact the interference level and the higher prioritized classes. One difference from SWAN is the approach to DiffServ; in SWAN the best-effort traffic is rate controlled, while the best-effort traffic is scheduled in the lowest priority class in this work. The latter approach is more in compliance with the differentiation idea of DiffServ, and consequently the transferability to DiffServ in fixed networks.

A QoS aware MAC layer would be an addition to the proposed QoS architecture. Admission control and preemption are still needed on higher layers. However, the presence of a QoS aware MAC layer would make the differentiation more clear, and would provide better protection for the high priority traffic, included routing and other control messages. QoS routing, on the other hand, provides resource allocation along the path during the route discovery phase. QoS routing includes a control for each hop along the path, while this work uses a source-based admission control, which is analogous to DiffServ. In DiffServ the admission is done at the edge, while the core routers only perform forward operations with low complexity. Keeping the complexity down in intermediate nodes is therefore in line with the DiffServ idea, while it also reduces the computations and power consumption overall in the MANET.

3. Fairness

Fairness is to give a fair allocation of the resources to all requesting parties when the network is unable to handle the demand, where a fair allocation will depend on the network objective. Fairness is, and has been, considered as an important property in most communication systems, and particularly in distributed systems where the available resources are to be shared by a number of users [JCH84]. There are several situations that can cause unfairness; among these are the existence of multiple bit rates within the same network which leads to uneven channel times [JL05], and different signal strengths in use that may cause a physical channel capture [GRG⁺06]. In this thesis, the focus on fairness is on the predictability of performance; a service needs some degree of predictability to be a viable service for users. Predictability, along with classification and quantification of fairness, is further discussed in this section.

Predictability

The traffic pattern and node distribution in a MANET can lead to a performance that is dependent on location; simultaneous transmitting neighbors, or a receiver located

multiple hops away, are aspects that can decrease the experienced performance. From the user's point of view, such aspects are invisible and will only lead to a less predictable service. Predictability is a very important factor in the use of communication services. Unpredictable service performances will lower the user acceptance, not only for the services themselves, but also for the network as a whole.

Location independent performance has been studied to increase the predictability. Lowering the local congestion level has been approached through the routing process by Giovanardi and Mazzini [GM07], where they increase the cost of using congested links in the route discovery phase. Multi-hop communication will be dependent on the locations of source and receiver; contention resolution at each hop leads to more retransmissions, an increased delay and loss probability, and lower throughput. Gerla et al. [GTB99] found that with TCP over a CSMA MAC layer, the throughput collapsed for paths over two hops. This problem has been approached in a proposal by Hsieh and Sivakumar [HS01]. They reduce the average number of flows sharing a link in the routing process to lower the probability of having flows with different path lengths on the same link. In another proposal by Jun and Sichitiu [JS03], the originating traffic was isolated from the relayed traffic. They discussed the generic problem of bottleneck links, with focus on the per-flow throughput ratio for different queue weights. Through simulations with both the standard IEEE 802.11 MAC layer and a QoS enabled MAC layer, they were able to achieve an improved fairness.

The idea of differentiating originating and relayed traffic is followed up in Paper D. However, Paper D approaches the problem of predictability in a more complete DiffServ context, where also the relations between the DiffServ classes are investigated. In addition, different policies related to the path information are tried, where one policy is to differentiate originating traffic from relayed traffic.

An uneven distribution of the per-node throughput will in many cases lead to a higher total throughput in the network [JL05, GTB99]. Communicating parties with good conditions, for example a one-hop path or a path with minimal interference, can be assigned the channel access to increase the total throughput; giving channel time to all nodes introduces a coordination overhead. The trade-off between fairness and utilization means that fairness in MANETs will not come without a cost, which is also observed in Paper D.

Classification of Fairness

Fairness has different definitions and can be achieved at different levels and time scales. Two ways of defining fairness is *min-max fairness* and *proportional fairness* [JS03]. Under min-max fairness, all nodes are split in two groups. Nodes in the first group, which consist of nodes that do not meet their QoS requirements, receive the same share of bandwidth. The other group consists of nodes that need less bandwidth than their fair share. Proportional fairness considers the nodes' offered load and maintains the nodes' throughput in proportions to their offered load. A node with one Constant Bit-Rate (CBR) flow gets half the throughput of a node with two such CBR flows. Jiang and Liew [JL05] discuss how proportional fairness in ad hoc networks

Table 2. Statistical indicators of variation

average	$\mu = \frac{1}{n} \cdot \sum_{i=1}^n x_i$
sample variance	$\sigma^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n x_i - \mu$
min-max ratio	$\frac{\min_n\{x_i\}}{\max_n\{x_j\}} = \min_n \left\{ \frac{x_i}{x_j} \right\}, \forall i, j = 1, \dots, n$
CoV of the sample data	$\frac{\sigma}{\mu}$
Jain's Fairness Index (JFI)	$\frac{(\sum_{i=1}^n x_i)^2}{n \cdot \sum_{i=1}^n x_i^2}$

can be achieved. Another view is *per-node fairness*, which gives nodes equal access to the channel independent of the flows traversing it.

Short- and long-term fairness consider the fairness with respect to time. It does not need to be any relations between the time scales; a network with long-term fairness does not need to be fair on a short term. For a statistical best-effort channel access, the probability will in the long run converge to $1/N$ for N competing sources. However, on shorter time scales, the probability may differ from $1/N$.

Quantifying Fairness

Quantifying fairness is basically to give an indication of the variability in performance among observed objects. Jain et al. [JCH84] list four properties for a fairness measure: population size independence, scale and metric independence, boundedness, and continuity. Table 2 lists some statistical measures that are all indicators of variations, except for the average; for all equations, the number of observations is n . The sample variance, min-max ratio, and the sample data's Coefficient of Variation (CoV) are all independent of population size, but the sample variance is dependent on the scale of the observations. The CoV is not bounded, but comparison of CoV values will indicate the system with most fairness. The min-max ratio is bounded, but does not reflect the distribution of the observations between the minimum and maximum values.

Based on the limitations mentioned above, Jain et al. [JCH84] propose the Jain's Fairness Index (JFI) that meets all of the four defined properties. The JFI is also defined in Table 2. It is bounded between 1.0 and $1/n$. The analyzed system is regarded fair for 1.0, while the system is regarded unfair for $1/n$.

The variance, CoV and JFI picture the distribution of the observations. For example, if 9 out of 10 nodes get a throughput close to the average, the indicated fairness would be high even if the last node receives no throughput at all. On the other hand, the min-max ratio tells how the worst performing node performs compared to the best. In the example mentioned, it would therefore have indicated an unfair system. Different fairness indicators should be used based on what is most important; whether it is to

give equal performance for most of the nodes, or to ensure a minimum of performance for every node.

4. Research Methodology

This thesis is based on the work conducted in the included papers. The work follows a traditional research methodology consisting of three stages: formulating a *work hypothesis*, *hypothesis testing*, and *result validation*. The formulations of the hypotheses are derived through a series of design discussions, and have been described to target the goals for this thesis.

In the testing phase, the hypotheses have been evaluated by simulations. The reasons for using simulations as means to illustrate the proposed methods and features, are mainly the control and flexibility it offers. It is easy to define "what-if" scenarios to see both direct and indirect effects. Controlling the environmental parameters gives equal conditions when comparing the obtained results. The simulator used is the J-Sim network simulator with the wireless extension package [Hun]. J-Sim is an open source Java [Sun] based simulator, which can operate on any platform.

In the final stage, the results are validated. Most software contain errors, so both the simulator and new solutions must be checked for implementation errors and logical misbehavior in the validation process. The validation process consists of different kinds of testing, tracing, and feasibility analysis of the results. Analyzing and testing small system components and fractions of code before the parts are combined, and tracing single packets through the protocol layers and network, have given valuable information on the behavior and correctness. Parts have also been validated through comparison with other approaches in the research. All simulation scenarios were tested with 10 replications, while some scenarios were tested with even more replications.

Two alternatives, or supplements, to simulation for the QoS architecture are analytical modeling and real experiments. Analytical modeling of the QoS architecture requires a simplified model and scenario, something which also induces results with less reality. However, the use of well approved equations and expressions can give more depth and credibility to the results by validating the simulations, and make limit values more obtainable. Extending the QoS architecture to a real experiment will clearly give improved reality. However, a cost-benefit trade-off is involved with real experiments; the scalability is lower, and a small change in the simulation parameters may induce a high cost in a real set-up.

5. Thesis Context

The proposed QoS architecture in this thesis is defined under the DiffServ architecture. The QoS architecture provides classification of the traffic into pre-defined Behavior Aggregates (BAs), or *service classes*, and gives differentiated treatment through different scheduling priorities. Based on the work in Paper A, the number of service classes is set to three to ensure separation since aggressive traffic is present. The classes are labeled EF, AF, and Best-Effort (BE) according to the pre-defined BAs in DiffServ. Based on the QoS requirements, a traffic flow is classified into one

of the three service classes, which gives the PHB throughout the DiffServ domain (here: the MANET).

The ITU-T defines a model for multimedia QoS from an end-user viewpoint in their recommendation G.1010 [ITU01]. The model is intended to form a basis for defining realistic QoS classes in networks and associated QoS control mechanisms, and is used in this work to define the traffic flows' QoS requirements. Traffic conditioning for ensuring compliance to traffic profiles is not specifically addressed in this work. The flows in the network are not defined in terms of traffic profiles, but the traffic conditioning is to some extent handled by the admission control.

It is assumed that the observed MANET corresponds to a single DiffServ domain with a common policy; within the MANET, a common description of the BAs is deployed, giving packets the same PHB independent of node. In fixed networks the domain boundaries are clear and defined. In MANETs, the domain boundaries are dynamic, and this makes it more difficult to provide a common policy throughout the MANET at all times. It is assumed that a mechanism for policy information exchange exists within the MANET. How to handle SLAs between domains is out of scope for this thesis. Inter-domain activities include how to mark inbound traffic, and to ensure that data flows stay within their traffic profiles. Studies concerning the interconnection is one direction for further research.

The QoS architecture is decoupled from the routing protocol, which means that the operations are not dependent on any particular class of routing protocols. It is an objective of this work that the QoS architecture is valid in any context, independent of choice for connectivity. The results derived from this work have been based on the AODV routing protocol. From the QoS viewpoint taken in this work, the routing protocol should only provide the connectivity. AODV was mainly chosen due to the in-frequent changes in traffic pattern, and availability. But a reactive approach may cause high delays when packets are re-routed in case of broken links, and it was observed that route repairs in many cases gave a substantial packet delay. To increase the successful delivery of routing packets, the routing packets were given highest priority, i.e. scheduled in the EF class.

On the MAC level, the IEEE 802.11 standard is used to access the channel. It provides a best-effort service by treating all packets equally, independent of the PHB on the network layer. This is clearly a limitation since low priority traffic flows will affect high priority traffic flows. It is believed that a MAC layer with the capabilities of differentiating between data flows, like the IEEE 802.11e, would improve the differentiation among the classes. But even with a QoS aware MAC layer, there is a need for traffic control mechanisms and differentiation at the network layer.

6. Contributions

This thesis defines a QoS architecture for MANETs based on the DiffServ architecture. The starting point was to investigate whether the DiffServ architecture could be extended into MANETs for a more seamless integration with DiffServ enabled fixed networks. By working with this idea, necessary components for a QoS architecture were identified; an admission control of flows, and preemption in case of congestion,

were needed for protection of prioritized traffic. Also, differences in performance lead to a path dependent scheduling for improvement in predictability.

The DiffServ specification defines a number of different classes to meet different user requirements. But the characteristics of multi-hop wireless networks do not give the same conditions as the wired counterparts. Therefore, an analysis on how many classes the MANET can support and separate was performed. Here, separation is when two classes can be distinguished by different loss, delay, etc. The simulations were done for a simplified DiffServ architecture with strict priority schedulers, and different traffic types. Four classes can be supported with only well-behaved traffic, while with more unpredictable and aggressive traffic present there is a need for explicit control of the traffic. This indicates that classes must be re-marked between the MANET and the fixed network; maybe running all AF class into a single AF class. Others have also used a low number of classes in MANETs, but have not followed the same rigid approach to find a number as in this thesis.

DiffServ needs admission control to protect existing traffic. An admission control based on shadow classes is proposed for the QoS architecture. Every flow with QoS requirements must be tried in the network before inclusion. A class will have an associated shadow class that is differentiated through higher drop probabilities in the buffers. This minimizes the impact new traffic will have on existing traffic from the same class. If admitted, the new flow is "upgraded" and sent in its designated class. Figure 12 illustrates the shadow class concept. The source node marks packets from a new flow, and each intermediate node monitors the buffer size. The receiver replies with a QoS report at periodic time intervals. Compared to other approaches as for example SWAN and CLAD, this scheme uses the real data flow during the admission to really see all impacts the flow will have, and thus avoids the bandwidth estimation problem. Notice that the IEEE 802.11 MAC is used, and since this MAC protocol can only provide a best-effort channel access, the resulting QoS may not be as evident as it seems on the network layer. The reasons for using the IEEE 802.11 standard is to better identify effects of the QoS mechanisms on higher layers. The QoS architecture is decoupled from the routing process in order to be independent of the routing protocol.

Due to topology changes with subsequent capacity and traffic pattern variations, a congestion control for the traffic is needed. The solution is to indicate congestion by the ECN mechanism. The RED mechanism monitors the buffers, and for high buffer sizes it either drops a packet or marks the packets through the ECN bits. The policy on what to do is decided by the flow's status (admitted or not). To give the admitted flows better protection, the probability of dropping packets from non-admitted flows is higher than marking packets from admitted flows for the same buffer size, as Figure 12 shows. Both the admission and congestion control will then be based on the same probabilistic RED functionality for control of the traffic. The importance of the RED parameter setting necessitated a sensitivity analysis. However, the analysis was simplified due to the complex relations that exists between the classes, which make it hard to identify all effects.

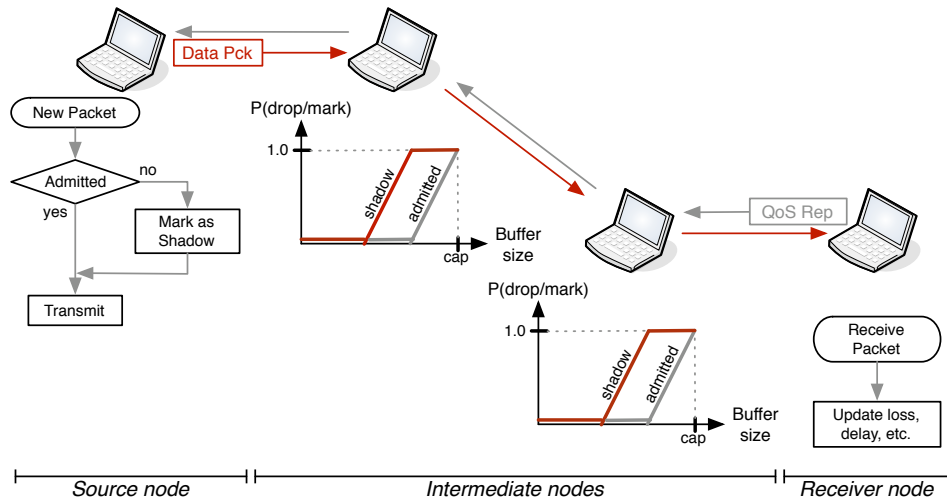


Figure 12. The shadow class concept

For a predefined number of ECN marked packets, a source will preempt its flow. Alternatives to preemption is source adaptations, or priority and/or service degradation. However, it is difficult for a source to know how to adapt to the new conditions, and a degradation of the flow will still affect other traffic due to the best-effort MAC protocol. Preemption, and a consequent re-probing, is also used in the SWAN architecture. Proposals using the IntServ approach, like INSIGNIA, mostly deploy degraded service or source adaptations in case the requested service can not be supported. But as stated, a degraded flow, e.g. running in best-effort mode, would still affect other flows and contribute to enforce the congestion already present.

The total throughput in the network can be maximized by giving the channel to nodes with the best conditions. However, fairness is important, and should be targeted to give all users a more predictable and even performance. To improve users' predictability of a service that is provided over a given path, a path dependent scheduling is included in the QoS architecture. Two schemes are proposed and investigated; differentiating on the number of hops made, and differentiating on the number of hops remaining. The idea of separating originating and relayed traffic has been approached by Jun and Sichitiu [JS03]. This approach studied a generic problem with bottleneck links, while the path dependent scheduling approaches the problem in a more complete DiffServ context. It also investigates the relations between the classes used, and for two different schemes. If fairness is important, the path dependent scheduling can be a viable solution, but at the cost of utilization and to some degree QoS. A challenge is to add more dynamics in the path dependent scheduling, which would involve the resource allocation, and how it could vary with time.

To sum up, Paper A starts to extend the DiffServ architecture into MANETs, and investigate the number of classes supported. The QoS architecture is proposed in Paper B with admission control based on shadow classes, before ECN with preemption is added in Paper C. A sensitivity analysis on the admission control used in Papers B and C is provided in Paper E. From observations in Papers A-C, the problem of unpredictability is approached in Paper D. In the next section, each paper is discussed in more detail.

6.1 Investigations and Contributions by Paper

Paper A

DiffServ in Ad Hoc Networks

Ad hoc networks will necessarily have a need for communication with fixed networks, for example access to the Internet. Also, there is a need for QoS between the two domains. Having different solutions for QoS provisioning in the two domains may clearly give interoperable problems. This was the motivation for investigating an extension of the DiffServ architecture into ad hoc networks.

Different approaches to DiffServ have been studied in the context of ad hoc networks in the literature. However, the number of classes in use has not been thoroughly investigated. This paper investigates the number of classes that the ad hoc network can support while still maintaining separation between them. The classes are differentiated by a strict priority scheduler (without preemption). The effects are investigated both with well-behaved CBR traffic, and more unpredictable aggressive traffic like TCP.

With only well-behaved CBR traffic in the network, the number of classes with sufficient separation is four. However, with TCP in the lowest priority class, the number of classes decreases to two. Three conclusions can be derived from this. First, the number of classes is, as expected, dependent on the types of traffic in the network. Second, for a specified number of classes, explicit control mechanisms are needed to maintain the separation when more aggressive traffic is used. And third, the results show that ad hoc networks are not capable of accommodating the same number of classes with separation as fixed networks. The study is quite simple, but serves to illustrate the constrained network conditions. The next step, which is pursued in Paper B, is to include admission control for controlling the offered traffic and protecting the network's existing traffic.

Paper B

QoS in Ad Hoc Networks

The paper extends the analysis of the previous paper. The augmented DiffServ architecture proposed in this paper, includes admission control to meet the goal of QoS provisioning. Based on Paper A, three service classes are used, with both CBR and

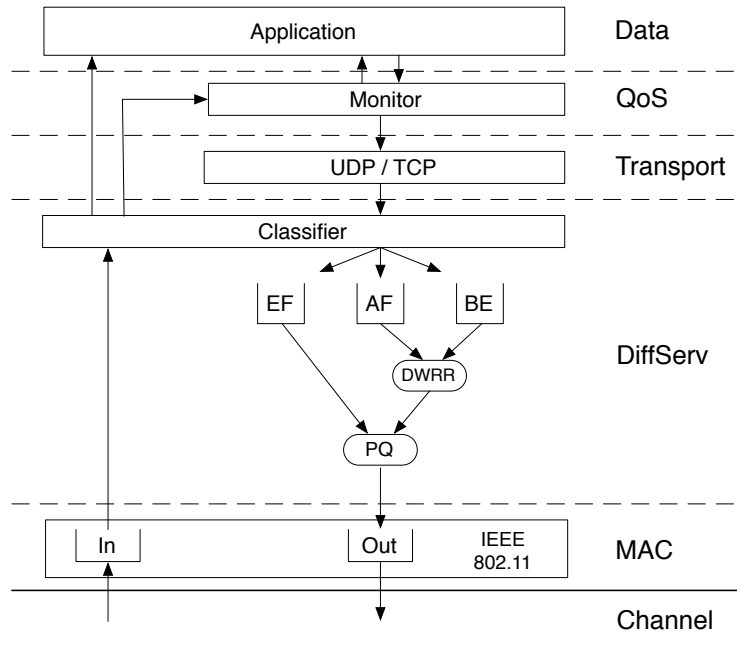


Figure 13. The layered approach of the QoS architecture

TCP traffic in the network. The architecture adds a QoS layer containing a monitor, which supervises each data flow. Figure 13 shows the layered architecture.

The proposed admission control is based on the concept of shadow classes. Many approaches, like SWAN and CLAD, use explicit probing to check a path's conditions. The shadow class approach will instead use the real data flow to see whether the network can accommodate it or not. When a new flow is started, it is sent in its class' associated shadow class instead of its originally designated class. The flows in the shadow class are differentiated from admitted flows by higher drop probability in the buffers through the RED mechanism. The higher drop probability protects the existing traffic from QoS degradation. After the admission phase, the monitor compares the QoS requirements with reports from the receiver and admits the flow or not. Figure 12 illustrated the shadow class concept.

From the simulation results, the high priority EF class experiences relative stable loss and delay for increasing traffic load, while the AF class gets a constant throughput when admitted. Also, the BE class, which is run without any admission control due to no QoS requirements, manages to utilize the available resources and get a high throughput.

The proposed admission control shows to be a viable solution for a DiffServ supported MANET. The mechanism has low complexity, and involves a low computational overhead for intermediate nodes. There might be moments of instability in the system when there are many shadow class flows waiting for admission. However,

the shadow class concept is a statistical call admission scheme, which aims at a good average behavior, that may fall short in certain scenarios.

Paper C

QoS Architecture in Ad Hoc Networks: Effects of Shadow Classes and ECN to Regulate the Load

From the previous paper, an admission control was proposed to the QoS architecture. It controls the offered load and protects admitted traffic. But due to the dynamic topology, congestion may appear.

In this paper, the QoS architecture is extended with the ECN mechanism. The RED mechanism used for differentiating the original classes from their associated shadow classes, is here also used for marking packets. In the IP-header, the ECN-bits are set to indicate congestion. It is difficult for a source to adapt correctly to the notification, and a degradation of a flow to for example the BE class will still affect existing traffic and the congestion. After a predefined number of ECN marked packets, the source therefore preempts its flow.

By incorporating ECN through the RED mechanism, the QoS architecture includes two traffic control mechanisms based on the same probabilistic function in the buffers. With the presence of ECN, also the TCP traffic is controlled by other means than the window mechanisms. To further increase the protection of traffic with QoS requirements, the marking of BE packets is dependent on the traffic volume in the higher priority classes in addition to the BE traffic. The EF and AF sources emulate voice and video applications, respectively, while TCP traffic is sent in the BE class. The admission control parameters are set according to recommendations by ITU-T for voice and video.

The admission control and congestion control are analyzed both individually and together for different strictness in the network. As stronger control is deployed, the ad hoc network is able to meet the applications' QoS requirements. However, with stronger control the utilization of the resources is decreased, i.e. the total throughput is decreased. The analysis shows that by adjusting the traffic control mechanisms individually, the network can be adapted to different network scenarios and goals of operation.

Paper D

Evaluation of Path Dependent Scheduling in Ad Hoc Networks: a Suitable Fairness Mechanism?

Fairness is an important property in communication systems, and observations from Papers A-C showed that the node's performances were unbalanced; some nodes seize most of the resources at the cost of the other nodes in the network. One underlying problem is that the performance drops with the number of hops in the MANET. The overhead for sharing a channel among several nodes, and the fact that packets along a

multi-hop path interfere with each other, give incentives for assigning the resources to the most effective nodes. Something which will increase the total throughput.

In this paper, the objective is a more even performance for increased predictability. Increasing the predictability gives more equal user expectations independent of path length used. The predictability is targeted by a path dependent scheduling, in where the packets' scheduling priority within a class is based on their path information. Two schemes are investigated. The first scheme differentiates local packets from transfer packets (i.e. hops made), while the second scheme differentiates packets on their last hop from the other packets (i.e. hops remaining). The scheduling opens for additional levels of differentiation as well, but only two levels are tried here.

To get a better understanding of the effects, a simple static chain of wireless nodes is tried before moving into random topologies with mobility and full QoS architecture. For the static scenario, the TCP goodput is used as an indicator of multiple performance indicators like loss and delay, while both CBR and TCP traffic are used in the random topologies scenario.

Through simulations it is found that the path dependent scheduling is a potential mechanism to achieve better fairness among the nodes in the network. There are complex relations between classes in wireless networks, where a change in the scheduling at one node can affect many other nodes as well. In the static topology, the effect of more even performance is lower utilization and reduced maximum performance. In random topologies, the cost is a reduced absolute performance for the high priority EF and AF classes.

In Papers B and C, the trade-off is between QoS and utilization. Here, the trade-off is between fairness and utilization, and also to some degree between fairness and QoS. The weights (i.e. resource allocation) were varied for the two schemes, and this showed that varying the weights can meet various operational goals in the network. However, there is complexity involved with the parameter setting in a complete QoS architecture scenario. It is hard to isolate each individual effect in such a scenario. More optimal parameter settings could clearly be found. But the mechanisms need to be efficient over a large set of topologies and traffic loads, and the more the parameters are tuned, the less usable is the architecture.

Paper E

QoS Architecture in Ad Hoc Networks: Sensitivity Analysis of the Admission Control

In Paper B, the shadow class concept for use in the admission control is introduced. Flows in a shadow class are differentiated from admitted flows by higher drop probabilities in the buffers through the RED mechanism. The probability of losing a packet in a buffer depends on the traffic volume and the thresholds set in the RED mechanism. Also, the congestion control is managed by the RED mechanism. In Papers B and C, the thresholds were set basically to protect the admitted traffic by reserving most of the buffers for this traffic. However, a thorough analysis was not given.

This paper makes an effort to fill the gap on the missing analysis in the previous papers. It investigates the performance's sensitivity on the thresholds. Due to such an analysis' complexity, the admission control's RED threshold is analyzed for the highest priority EF class. The AF class depends on the EF class through the strict priority scheduler, while the BE class depends on the traffic volume in both of the other two classes. Although a not very realistic model, the M/M/1 queueing model gave clear indications of a too low threshold for the traffic load targeted. Therefore, the admission control is in this paper tested for a relaxed setting by higher thresholds, along with lower buffer capacity.

The analysis shows that the EF class is not very sensitive to the changes made, but very complex relations were discovered. All the classes will in a way depend on each other due to the shared wireless channel, in addition to the direct dependencies from the buffer control. The changes give large variations in the amount of delivered traffic from the lower classes (i.e. AF and BE), but the variation is explained through an analysis of the AF throughput.

To conclude, it can be seen that the direct effect of parameter tuning can be quite coarse-grained in ad hoc networks, and that each class' performance depends on the general traffic volume along with the scheduling made by all nodes within range. Compared to fixed networks, the ad hoc network can not provide the same granularity in effects for e.g. changes in buffer management or scheduling; the changing topology and dependence on all nodes within range make the conditions different from the fixed counterpart. This paper simplified the sensitivity analysis, and a more comprehensive analysis around each class' thresholds and their interrelation is for future research.

7. Conclusions

The QoS architecture defined in this thesis is defined under the DiffServ architecture for better integration with fixed DiffServ networks. It includes three main components: admission control, congestion control with preemption, and path dependent scheduling for improved predictability.

Providing differentiated treatment to the service classes in a MANET seems to be the best solution in the trade-off between scalability, overhead, and performance. The shadow class concept in the proposed admission control, which is distributed and source-based, admits flows on statistical measures, without any need for explicit bandwidth estimations or reservations. The only overhead is the feedback messages from the receiver, which can be piggybacked on data packets. Sending the data flow itself to test the network is basically the only way to see the real effects of a new flow. However, there are open issues around the shadow class concept. First of all, there is a trade-off between a short admission phase with few packets and a long phase with more packets. The former provides a quick response time, while the latter is more reliable. Then, since the admission control operates with good statistical performances in mind, there will always appear special cases, which need more attention.

Due to the dynamic topology, congestion control is needed. The same statistically mechanism is used to avoid congestion as to drop packets from shadow classes. ECN marking through the RED mechanism is a well proven method, and seems to work

well in ad hoc networks. Issues here are how the source should react, and how the performance is affected by lost congestion indications. For a source, it is difficult to correctly adapt to congestion notifications, and decreasing a flow's priority (i.e. service degradation) will not ease the congestion. Therefore, preemption is used in case of congestion.

An important property in communication systems is fairness. In multi-hop environments, the performance will be location and path dependent. A method based on path information is tried on the QoS architecture to increase the predictable performance for a given path. The path dependent scheduling is able to improve the predictability by assigning resources depending on the path length from source to receiver. The downside is a lower call admission ratio and lower absolute performance for the real-time traffic.

Based on the findings in the included papers, the QoS architecture, with shadow classes and ECN, is a potential mechanism to provide QoS in ad hoc networks. Also, the differentiation of traffic based on path information is shown to improve the predictability. However, both QoS provisioning and fairness are trade-offs between the utilization of network resources (i.e. total throughput) and absolute performance.

Among the challenges and directions for further research is more work on the three main components in the QoS architecture: admission control, congestion control, and path dependent scheduling. Also, the inclusion of a QoS enabled MAC protocol would be a valuable amendment to the architecture. For example, the access categories classified by the IEEE 802.11e could be mapped to the QoS architecture's service classes, and to DSCPs in other DiffServ operated networks. Other challenges related to this thesis are investigations concerning the interconnection with fixed networks, and more dynamics to the path dependent scheduling. Investigations concerning the interconnection with fixed networks, and how any agreement between the domains should be designed, is of importance for the services provided in MANETs.

A solution to the many possible settings of the QoS architecture's three components, could be to define profiles according to some important objectives, which may suit a range of conditions experienced in a MANET. Defined profiles would make it easier to target a specific situation and scenario more precise and with correct means.

II

INCLUDED PAPERS

Paper A

DiffServ in Ad Hoc Networks

Tor K. Moseng and Øivind Kure

Lecture Notes in Computer Science (LNCS)

Vol. 4396/2007, May, 2007

PAPER A: DIFFSERV IN AD HOC NETWORKS

Tor K Moseng, Øivind Kure
Centre for Quantifiable Quality of Service in Communication Systems
Norwegian University of Science and Technology,
Trondheim, Norway
{torkjeti, okure}@q2s.ntnu.no

Abstract In this paper we study the expected difference between the QoS classes in an ad hoc network. The results have a direct bearing on the suitability of extending a fixed DiffServ architecture into an ad hoc network. Through simulation, we analyze the number of classes that can be used in the ad hoc network with separation between the observed QoS in the different classes. The results clearly depend on the type of traffic run in the network. With well behaved CBR traffic, the ad hoc network supports no more than four classes, but with more aggressive traffic like TCP no more than two classes are supported. In addition, there is a fairness problem; the performance for a particular flow is not well distributed among the nodes.

1. Introduction

In an ad hoc network the nodes communicate over a shared wireless channel. There is no infrastructure and the nodes must operate as a router to forward traffic to destinations that are multiple hops away. Ad hoc networks have therefore different characteristics compared to fixed networks. A common channel with hidden and exposed nodes combined with node movement, are aspects that necessitate different solutions and mechanisms. The dynamic nature of the links makes Quality of Service (QoS) provisioning more challenging than in comparable sized fixed networks.

One of the most promising QoS architectures for fixed networks is the DiffServ architecture [BBC⁺98]. DiffServ is a QoS architecture, which aggregates each flow into a defined class, and treats all packets on a per-hop basis based on which class the packets' flow belongs to. DiffServ solves the scalability problems in the IntServ architecture [BCS94]. IntServ is a flow-aware architecture that does not scale if the number of source-receiver pairs becomes too high. The DiffServ architecture uses a 6-bit Differentiated Service Code Point (DSCP), which is in the Type of Service (ToS) field in the IP-header, to define a packet's treatment in the network. This treatment gives the per-hop behaviour (PHB) in each intermediate node that belongs to the network domain. The domain's policies will therefore give the definite treatment. Packets receiving similar forwarding behaviour belong to the same PHB class - or

DiffServ class. The differentiation among the classes is typically given by the buffer capacity, buffer drop probability and packet scheduling technique. DiffServ defines one Expedited Forwarding (EF) class, twelve Assured Forwarding (AF) classes (i.e. four classes, each with three drop probabilities) and one Best-Effort (BE) class.

In this paper we explore the number of QoS classes that can be supported in an ad hoc network. Clearly an IntServ based architecture is feasible, but our work aims at exploring the number of classes ad hoc networks can sustain assuming a DiffServ based architecture. By sustaining we mean the number of classes where the flows' end-to-end performance in the various classes differs markedly. This will be the starting point in defining a suitable QoS architecture, or less drastically, suitably mapping schemes between the standardized DiffServ classes and the corresponding classes in ad hoc networks.

There are several QoS architectures proposed for ad hoc networks. The number of aggregation classes range from two classes [LC05], and up to four classes based on 3GPP's class definitions [rGPPG04, ZML05]. However none of the different proposals make a stringent argument for the number of aggregation classes selected. It is therefore beneficial to analyze the maximum number of classes in a network where there is still some separation between the observed behaviour of flows mapped to the different classes. The separation in behaviour is clearly a function of the per node behaviour for each class. We use the standardized per hop behaviour in DiffServ with one EF class, a set of AF classes and a BE class. The AF classes differ in terms of drop probability and in the fraction of the forwarding resources allocated to the classes.

Through simulations we find the separation in terms of the performance characteristics loss and delay for different number of aggregation classes. The degree of unpredictability in how the ad hoc network will influence the various traffic classes is likely to be a function of the type of traffic run in the aggregation classes. Well behaved CBR traffic is relatively easy to deal with compared to more aggressive traffic like TCP. We will therefore perform a sensitivity analysis for different types of traffic in the "lowest" aggregation class, which we call the BE class for simplicity.

QoS architectures for ad hoc networks in today's literature include proposals like INSIGNIA [LAZC00], ASAP [XSA03], FQMM [XSLC00] and SWAN [ACVS02]. INSIGNIA is an end-to-end resource reservation based protocol. ASAP extends INSIGNIA with hard and soft reservation modes to utilize the resources better. FQMM combines IntServ and DiffServ where only the highest prioritized classes may use IntServ because of its per-flow guarantee and lack of scalability. SWAN is a class based approach that uses probing for admission control of the real-time packets, and rate control for the best-effort packets. All these architectures are only intended for the ad hoc network and need mapping of some kind in gateways for Internet connectivity. There is some research that cover the interconnection point between a fixed and an ad hoc network [DR04a, DR04b, DR05, SK05, CSNV05, MK03]. However, these approaches require mapping between two different architectures in the gateways. References [DR04a, DR04b, DR05, SK05, CSNV05] extend SWAN to cope with the interconnection point, while [MK03] considers inter-domain agreements and proposes

a framework to help the ad hoc and fixed network to cooperate. Clearly, using the same architecture in both the fixed and ad hoc networks would be beneficial.

The rest of the paper is structured as follows: Section 2 describes the simulation setup. Section 3 and Section 4 give the results from the simulations. Section 5 presents a different view on the network by considering the fairness. And finally, the paper is concluded in Section 6.

2. Simulation Setup

The main goal with the simulations was to make a sensitivity analysis on the number of classes in the network. The number of sources were kept constant in our network, so in order to increase the number of classes in the network, we chose to split up the lowest aggregated class (here: the BE class) in more classes. Also, we wanted to see how the performance depended on the type of traffic in the network. We started therefore with only well behaved traffic like CBR traffic. We then introduced TCP as aggressive traffic in the lowest aggregated class, i.e. the BE class. The TCP sources was only restricted by their sending window that was set to 64 kB. The now more aggressive BE traffic increased the offered load to the network, compared to the scenario with only well behaved CBR sources. Also, the load was different for different number of classes since the number of BE nodes was varied. However, even though this can explain some of the results, the conclusions found in this paper are still valid.

Regarding the buffer scheduling and buffer management mechanisms, we wanted to keep it simple. To fully exploit mechanisms like WRR and RED, which are used in the standard DiffServ architecture, one need to do a sensitivity analysis on how to best configure them. This was evident when we did a few simulations with RED as our buffer management. With the same RED parameters for all priority classes, there was a non-existent separation among the CBR traffic classes. Based on this, we used Strict Priority and Drop Tail as our buffer scheduling and buffer management mechanisms, respectively. However, studies with different buffer scheduling and buffer management mechanisms will be a part of our future work.

The simulations were done using the J-sim network simulator, version 1.3 and patch 3, with the wireless extension package [Hun]. A 1500 x 300 meters area was used with a total of 40 nodes. 18 of these nodes were selected as a source and divided into a set of classes, varying from two to six. A random receiver was by each source elected from the remaining 22 nodes. The sources were divided into the following class distribution (EF:AF1:AF2:AF3:AF4:BE): (3:0:0:0:0:15) for two classes, (3:6:0:0:0:9) for three classes, (3:3:3:0:0:9) for four classes, (3:3:3:3:0:6) for five classes and (3:3:3:3:3:3) for six classes.

Each source sending CBR traffic, transmits with a fixed rate from 5 to 30 kbps according to the scenario, 10 replications - each with a different seed, running for 200 seconds. Each source started sending packets after a negative exponential distributed

waiting time to avoid the same starting time. The average packet size was 358 bytes¹. The interdeparture time varies according to a negative exponential distribution with mean equal to the data rate. Every node drew a new starting point for each repetition and moved within the simulation area according to the Random Waypoint model with a 2 m/s mobility. The IEEE 802.11 MAC-layer was used with 2 Mbps capacity and 250 meters transmission range. Each class' buffer size was computed from the buffer delay restriction placed on each of the classes: 3 kB (EF), 18 kB (AF1), 31 kB (AF2), 44 kB (AF3), 57 kB (AF4) and 70 kB (BE) were used. The reactive routing protocol AODV [PBRD03] was used as routing protocol in the simulations. The AODV buffer was set to 64 kB. By setting the RTS-threshold to zero, every packet was sent with an RTS/CTS handshake first. The AODV HELLO mechanism was disabled, which means that link breaks were discovered by retransmissions.

3. Results

To compare the results we chose loss and delay to illustrate the differentiation between the QoS in the classes. The loss is an average over all packets for each source in the aggregate, while the delay is an average for all received packets for each source in the aggregate. This gives a packet's view of the network and tells the probability of losing a packet and gives the expected delay for a packet. A different view, from a node's point of view, reflects the fairness in the network and is explained and discussed in Section 5.

As we expected, the packet loss in the EF class is independent of the number of classes in the network. This is shown in Figure 1. What is also shown is that the BE class's packet loss is independent of the number of classes in the network, and quite similar to the EF class's packet loss. One could have expected that the BE loss would increase for more classes in the network, but this effect seems to be minimal. Lower priority and larger buffer size gives EF and BE, respectively, nearly the same packet loss. Packet loss because of buffer overflow on the IP-level is not common when only around 10% of the dropped packets are due to buffer overflows. The high packet loss is caused by the high traffic intensity in the network and no existing route between source and receiver. The CBR sources will send at the same data rate independent of the network conditions and whether a valid route exists. All the noise on the shared channel leads to retransmissions and disrupts the packets, while no existing route leads to packet drops in the AODV buffer. This will consequently lead to a high loss rate. However, the loss rates will not change the paper's conclusions.

In terms of differentiation, the delay is different from the loss. Starting with two classes, the nodes receive the same treatment for low load, but when the load increases the difference between the EF and BE packets become more evident. From bit rates higher than 10 kbps, the EF packets are differentiated from the BE packets (Figure 2(a)). When we introduce another class in the network, the same pattern

¹The packet size was drawn from a distribution; 128 byte packet with a 0.2 probability, 256 byte packet with a 0.5 probability and 512 byte packet with a 0.3 probability.

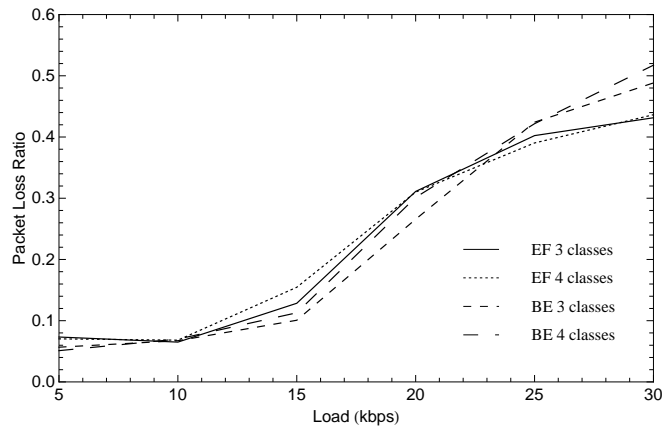


Figure 1. EF and BE loss with CBR BE traffic

appears, as shown in Figure 2(b). The only difference is that the behaviour is equal up to 15 kbps. For higher bit rates, the classes are differentiated. The EF packets get the lowest delay, followed by the AF1 packets and BE packets in that order. This is as expected because of the priority and the smaller buffer sizes for higher prioritized classes.

Increasing to four classes (Figure 2(c)) show the exact same pattern, where AF1 and AF2 are between the EF and BE classes - AF1 with lower delay than AF2. But when five classes are introduced, the differentiation gets a bit blurred, as Figure 2(d) shows. Again, the classes show different values from 10 kbps and above. However, the classes are not completely separated over the entire load range. The EF and AF1 classes cannot be separated from each other, and also, the AF2 and AF3 classes intercept at 20 kbps. Incrementing further with one class in the network blurs the pattern even more. So, to fully separate the classes from each other above the delay knee (i.e. the point where the classes start to differentiate), the number of classes must be kept below five. The long delays must be commented. The long delays observed in Figure 2 are insufficient for most flows. The reason for the long delays is the simulation setup that causes high traffic intensity in the network. However, the delay values are, as the loss rates in Figure 1, a function of the simulation setup and will not change the conclusions in this paper. With the setup, the EF and AF traffic are not close to an acceptable operational point. Still we compared the delay to illustrate the degree of separation even under heavy disturbance. Due to the heavy loss, the detailed perturbation of the curves is of less interest since the paths' packets are lost and will affect the average delay. Another issue is that there is a large variation on what delay the packets experience. The large buffer sizes combined with the traffic intensity produced some worst case delays for some packets, which is reflected in the high average delays in Figure 2.

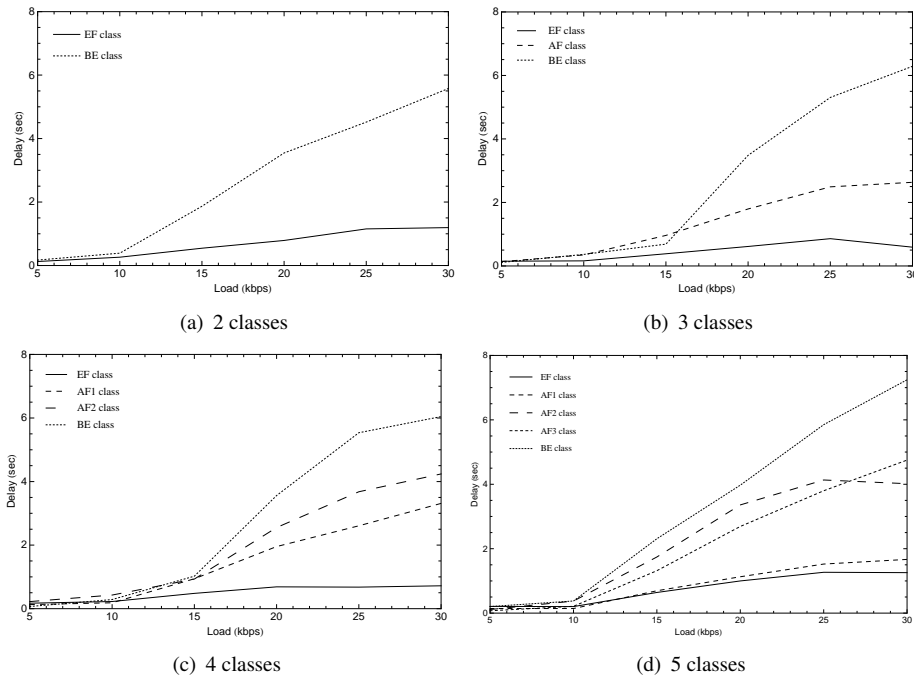


Figure 2. Differentiation in terms of delay with CBR BE traffic

4. Aggressive Best-Effort Traffic

For the first results, described in Section 3, all traffic was well behaving. Each packet was sent according to a constant bit rate scheme. However, more aggressive traffic like the Transport Control Protocol (TCP) [Inf81] are likely to affect the degree of separation between the classes. The simulation was therefore repeated with TCP traffic in the lowest aggregated class (i.e. the BE class). We used the TCP Reno version [Jac90] as it is a well deployed version. Our TCP sources had no bandwidth limits - they were only restricted by the sending window, which was set to 64 kB (no restrictions on the receiving window).

Figure 3(a) and 3(b) show the packet loss for three and four classes in the network, respectively, as a function of the load. The load on the x-axis is in this section for the nodes running CBR, while the BE nodes run TCP and are, as mentioned, only restricted by the sending window. As seen from Figure 3 the CBR nodes have a very high packet loss that is quite stable independent of the bit rate. As mentioned in Section 3, the reason for the high loss is because of the high traffic intensity in the network and no existing route. The traffic intensity on the channel is even higher here than in Section 3 when TCP was not present. The number of packets sent by a BE node can be over 10 times the number sent by an EF node for low bit rates. Opposed to the EF nodes, which send packets at fixed rate independent of the channel conditions and whether there exists any valid route to the receiver, the BE nodes throttle their sending rate down to zero if no acknowledgements are received or no valid route exists.

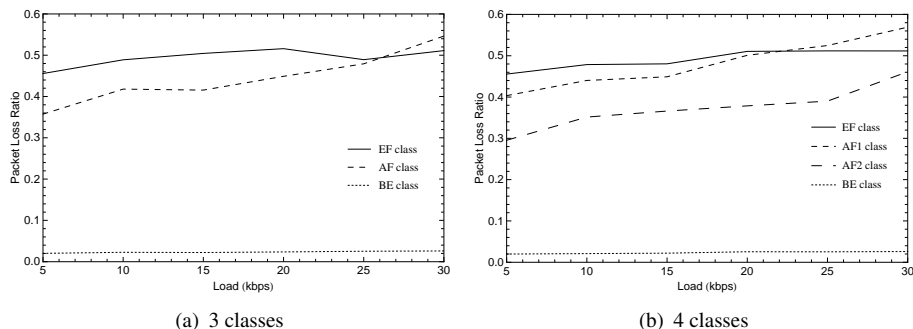


Figure 3. Packet loss rate with TCP BE traffic

However, if there exist a good path, TCP increases its sending window exponentially and the transmission rate is then also increased substantially. Since TCP only sends many packets when the path is good, the loss probability is much lower than what is experienced in the classes with CBR traffic. As we could expect, both AF1 and AF2 experience lower loss than the EF class because of the larger buffer sizes, with AF2 lower than AF1. This is true until the curves intercept at 25 kbps.

The packet loss for the EF class in the two classes scenario is higher than the loss in the three and four classes scenario. This is due to, as discussed previously, the reduction in the number of BE nodes from the two to three classes scenario (i.e. from 15 to 9 nodes). The same behaviour is seen for five and six classes where the loss is even lower - the number of BE nodes is for those cases further reduced. One could argue that the number of TCP sources should have remained constant for all the simulations. By decreasing the number of TCP sources as more classes are included, the EF class should be better off with more classes. However, as is shown in Figure 4, the separation is lost for more than two classes. This means that our conclusions still hold. A constant number of TCP sources, say nine TCP sources² for all number of classes, would only have given better differentiation with two classes compared to more classes in the network.

Considering the delay, two classes in the network (Figure 4(a)) show that the EF and BE nodes are differentiated for all bit rates. But the separation is already lost when three classes are in the network (Figure 4(b)). That is, there is separation between the classes, but the classes are not differentiated based on their relative importance since BE experiences lower delay than AF1. The same pattern also appears for four classes (Figure 4(c)). The introduction of aggressive traffic like TCP degrades the in-between classes (here: AF1 and AF2) relative to the aggressive traffic itself. As long as there is no MAC layer priority the access to the channel will be a function of the traffic load from the neighbours. The upcoming IEEE 802.11e [IEE07] standard aims to provide priority at the MAC level. This standard could be a possible enhancement in our

²9 TCP sources are used for three and four classes in the network. Using 9 TCP sources instead of 15 would therefore only have increased the separation with two classes in the network, which are already separated.

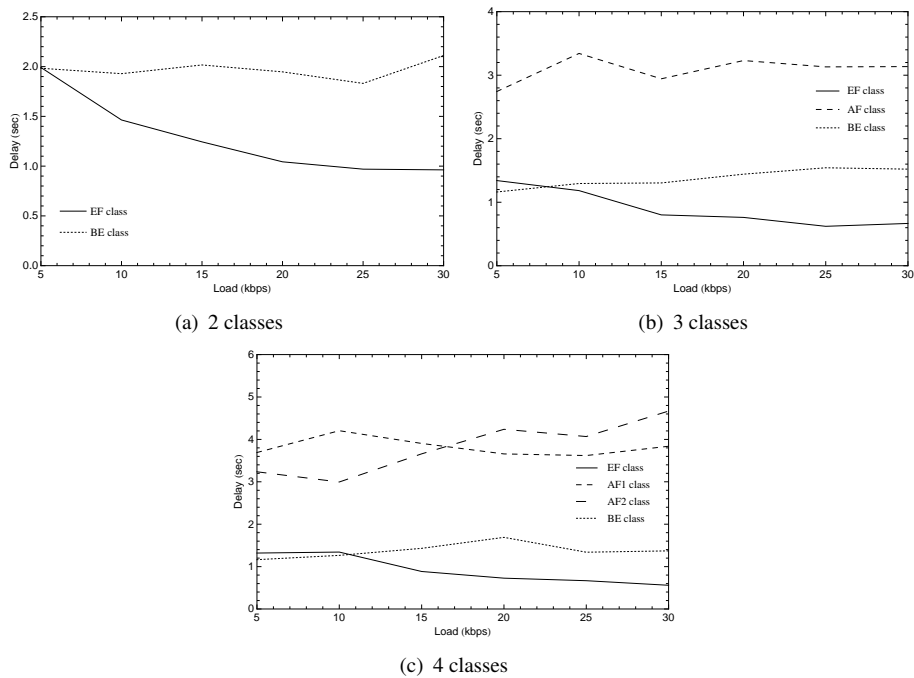


Figure 4. Differentiation in terms of delay with TCP BE traffic

model to strengthen the differentiation between the classes. The local IP scheduling differentiation only affects the packet handling on the node and not the channel access. For four classes the effect of packet scheduling is marginal compared to the overall delay. The delay is filtered through a high packet loss. The packet loss mostly affects the absolute level of delay, and not the relative performance. So, it is seen that the differentiation between the classes when nodes are transmitting aggressive traffic is lost already at three classes in the network. This is in contrast to the case where every node was well behaved in Section 3.

5. Fairness

The results that were presented in Section 3 and Section 4 were found from a packet's view of the network. From the packet's view the loss and delay values are averaged over all packets that are lost and received, respectively, by each node in an aggregated class for all repetitions. In this view, the packet has a certain probability for being lost and a certain delay it could expect, given the packet's class.

A different view of the network is from the node's point of view. The node's view considers the fairness aspect of the network. How is the loss and delay distributed among the nodes, and which loss and delay values can a new entering node expect? The aspect of fairness is a familiar problem in wireless conditions and is hard to deal with [CDL05]. In our simulations, the results given in this section are averaged over all nodes' individual loss and delay in an aggregated class for all repetitions. As an

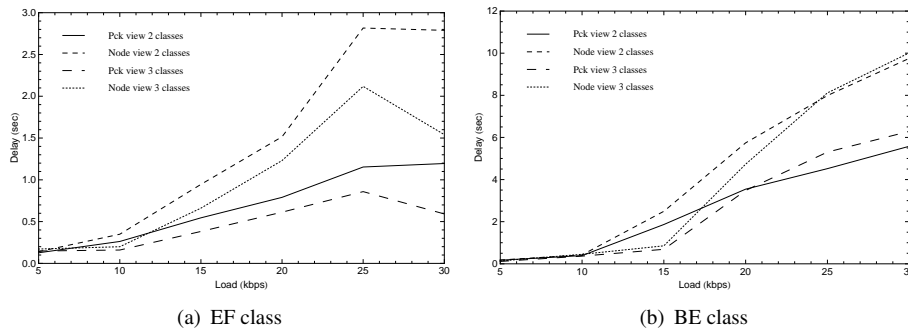


Figure 5. Delay with CBR BE traffic

example let us use one node with bad conditions where TCP restricts the number of packets to 500 with only 100 received i.e. a 0.8 loss probability, and a node with good conditions that sends 20000 packets of which 19900 packets are received, i.e. a 0.005 loss probability. Equal weight among these two nodes results in an average 0.4 loss probability, while from a packet's view there would be an average 0.01 loss probability. A new entering node will therefore expect a 0.4 loss probability in the network; even the network has a 0.01 loss probability as seen by the packet view.

With only well behaved CBR traffic the packet loss is only slightly different from the packet's view. Because the nodes are all transmitting nearly the same number of packets, and major packet drops like no valid route exist under both views, the loss do not differ much. The delay, on the other hand, shows more difference between the views. Nodes with only a few packets reaching the destination have generally higher average delays, and these will in the node's view get a higher weight and increase the class' average delay compared to the packet's view. So, if a node's view gives higher delay than the packet's view, it means that some nodes receive relative poor performance and will therefore increase the average delay since the nodes' weight is increased. This is seen in Figure 5(a) where the node's view doubles the average EF delay seen under the packet's view. Also, the average BE delay is substantially increased for higher bit rates - around 50% higher (Figure 5(b)). This shows fairness problems in the network.

Although the network gave fairness problems with only well behaved sources, introducing TCP only increases the unfairness in the network. Figure 6 presents the packet loss for the BE nodes to show the great difference between the two views. The packet loss is several orders higher in the node's view than in the packet's view, both for two and three classes in the network. Because a node with a bad route will get equal weight as a node with a very good route, as described above, the packet loss rate is shifted upwards. In the simulations a TCP node with a very good route may send 50 times more than a TCP node with a bad route in our simulations. As already stated: even though the packet loss rate in the network is low, the probability for packet losses for an individual node is much higher due to the shared wireless channel. Another illustration on the fairness problems is the loss distribution difference between the

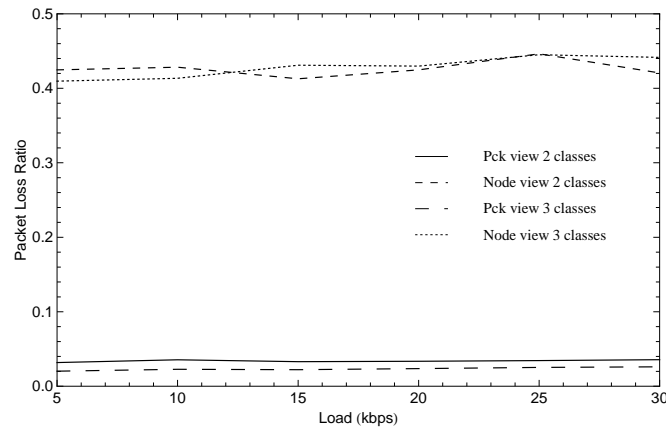


Figure 6. BE loss with TCP BE traffic

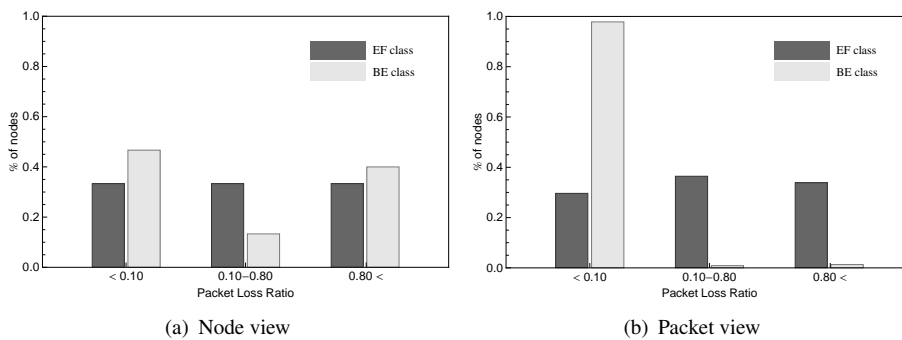


Figure 7. Loss distribution (TCP BE traffic)

packet's and the node's view. As an example, Figure 7 shows the loss distribution for one replication of the two classes scenario where the EF class sends 5 kbps CBR and the BE class sends TCP traffic. According to Figure 7(a) there are several nodes that experience a loss probability above 0.8, however according to the packet's view in Figure 7(b) almost all packets experience a loss probability below 0.1 in the network (this is especially true for TCP nodes). This shows that some nodes receive good performance at the expense of the rest of the nodes.

The same effect as the node's view has on the packet loss is also present for the delay. The delay for EF nodes is more than doubled for the two classes scenario, and around four times higher when running three classes in the network (see Figure 8). This implies that some nodes experience a very high delay that is spread over the total received packets in the aggregate for all replications under the packet's view, but is only spread over the number of nodes under the node's view, which increases the average delay. The delay is even worse for the TCP nodes (i.e. the BE class). This is because, as already mentioned, the big difference in the number of packets that are sent and received on a good versus a bad route. In addition, the high number of

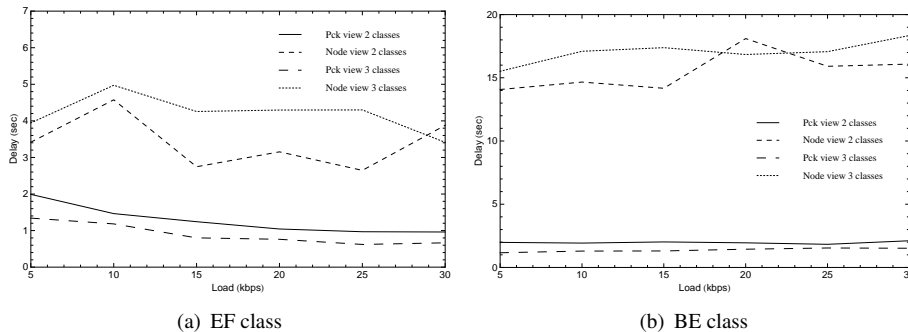


Figure 8. Delay distribution (TCP BE traffic)

packets received on a good route generally experiences a low delay, while the bad route generally experiences a high delay. So when each node gets the same weight, the average delay increases around 10 times dependent on the number of classes in the network (see Figure 8(b)).

We see that we need more explicit mechanisms to ensure fairness in the network. Nodes that control the channel transmit flows that have relatively few hops to the receiver, and we therefore need more global mechanisms rather than only local mechanisms in order to distribute the performance more evenly. A different buffer scheduler, e.g. WRR, could give a fairer share to the different queues in one node, but there are nodes that do not forward any traffic. If these nodes only transmit aggressive TCP traffic, they seem to control the shared channel as their own dedicated channel. Therefore, not only a fair buffer scheduler is needed, but also a mechanism that takes more global effects into count as regards network fairness.

6. Conclusions

The different context of wireless ad hoc networks compared to fixed networks necessitates different approaches for QoS provisioning. Several QoS architectures have been proposed for the ad hoc network independent of the fixed network. We investigate in this paper the suitability of using a class based QoS architecture in the ad hoc network. More precisely, we investigate the number of classes an ad hoc network can support while maintaining separation between the classes.

Based on simulations, a maximum of four classes are supported when the traffic is well behaved, i.e. CBR type traffic. Once more aggressive traffic sources like TCP are run in classes with large buffers, the ad hoc network is barely able to support two classes. Also, the highest prioritized classes, which send CBR traffic, experience a high loss because of the interference from the aggressive TCP traffic run in the lowest prioritized class. We must therefore shape the aggressive traffic in some way. By restricting the aggressive traffic, more QoS classes could be separated and other types of traffic would have a higher probability of accessing the shared channel.

The number of classes supported was also investigated from a fairness point of view. In the simulations, the performance was distributed unevenly among the nodes in the

network. Some nodes received good performance at the expense of others. There were nodes that could not get access to the channel because others were controlling the channel - especially with aggressive TCP flows. In order to ensure fairness for all flows independent of number of hops, we cannot rely only on local mechanisms - we also need some global mechanism.

For future work we will investigate suitable mechanisms for shaping the aggressive traffic in the network. Different mechanisms for the buffer management and scheduling will be a first step; some initial simulations showed that RED ensured a differentiation of TCP traffic from the higher prioritized CBR traffic quite well. However, a sensitivity analysis of the parameters will be required. Our analysis is aimed at ad hoc networks with no MAC layer QoS mechanisms. The new IEEE 802.11e MAC layer with up to four priority classes will clearly improve the support for a class based QoS architecture.

The simulations indicated that it is possible to use the default DiffServ architecture in ad hoc networks given some strong assumptions on the traffic behaviour; aggressive traffic need to be shaped in order to ensure timely access for higher prioritized traffic. Secondly, there need to be additional mechanisms to ensure fairness between the nodes in the network.

References

- [ACVS02] G.-S. Ahn, A.T. Campbell, A. Veres, and L.-H. Sun. SWAN: Service Differentiation in Stateless Wireless Ad Hoc Networks. In *Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, volume 2, pages 457–466, June 2002.
- [BBC⁺98] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. An Architecture for Differentiated Services. IETF RFC 2475, December 1998.
- [BCS94] R. Braden, D. Clark, and S. Shenker. Integrated Services in the Internet Architecture: an Overview. IETF RFC 1633, June 1994.
- [CDL05] C. Chaudet, D. Dhoutant, and I.G. Lassous. Performance issues with IEEE 802.11 in ad hoc networking. *IEEE Communications Magazine*, 43(7):110–116, July 2005.
- [CSNV05] S. Crisóstomo, S. Sargento, M. Natkaniec, and N. Vicari. A QoS Architecture Integrating Mobile Ad-Hoc and Infrastructure Networks. In *Proceedings of the 3rd ACS/IEEE International Conference on Computer Systems and Applications*, pages 897–903, 2005.
- [DR04a] M.C. Domingo and D. Remondo. A Cooperation Model between Ad Hoc Networks and Fixed Networks for Service Differentiation. In *Proceedings of 29th Annual IEEE International Conference on Local Computer Networks*, pages 692–693, November 2004.
- [DR04b] M.C. Domingo and D. Remondo. Analysis of VBR VoIP Traffic for Ad Hoc Connectivity with a Fixed IP Network. In *Proceedings of IEEE Vehicular Technology Conference (VTC)*, volume 4, pages 2834–2837, September 2004.
- [DR05] M.C. Domingo and D. Remondo. A Cooperation Model and Routing Protocol for QoS Support in Ad Hoc Networks Connected to Fixed IP Networks. In *Proceedings of Advanced Industrial Conference on Telecommunications/Service Assurance with Partial and Intermittent Resources Conference/E-Learning on Telecommunications Workshop (AICT/SAPIR/ELETE)*, pages 390–395, July 2005.
- [Hun] Hung-ying Tyan et al. J-sim network simulator. [Online] Available: <http://sites.google.com/site/jsimofficial/>.

- [IEE07] IEEE Computer Society. IEEE Standard for Information Technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications (IEEE Std 802.11-2007). IEEE Standard, [Online] Available: <http://www.ieee802.org/11/>, March 2007.
- [Inf81] Information Sciences Institute at USC, for DARPA. Transmission Control Protocol. IETF RFC 793, September 1981.
- [Jac90] V. Jacobson. Modified TCP Congestion Avoidance Algorithm. Technical Report. Email to the end2end-interest mailing list. [Online] Available: <ftp://ftp.isi.edu/end2end/end2end-interest-1990.mail>, April 1990.
- [LAZC00] S.-B. Lee, G.-S. Ahn, X. Zhang, and A.T. Campbell. INSIGNIA: An IP-Based Quality of Service Framework for Mobile Ad Hoc Networks. *Journal of Parallel and Distributed Computing, Special Issue on Wireless and Mobile Computing and Communications*, 60(4):374–406, April 2000.
- [LC05] X. Li and L. Cuthbert. Multipath QoS Routing of supporting DiffServ in Mobile Ad hoc Networks. In *Proceedings of the ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD/SAWN)*, pages 308–313, May 2005.
- [MK03] Y.L. Morgan and T. Kunz. PYLON: An Architectural Framework for Ad-Hoc QoS Interconnectivity with Access Domains. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences (HICSS)*, January 2003.
- [PBRD03] C. Perkins, E. Belding-Royer, and S. Das. Ad Hoc On-Demand Distance Vector (AODV) Routing. IETF RFC 3561, July 2003.
- [rGPPG04] 3rd Generation Partnership Project (3GPP). Quality of Service (QoS) Concept and Architecture (Release 6). Technical Specification Group Services and System Aspects, TS23.107 v6.1.0, March 2004.
- [SK05] C. ShanShan and A.J. Kassler. Extending SWAN to Provide QoS for MANETs Connected to the Internet. In *Proceedings of International Symposium on Wireless Communication Systems (ISWCS)*, pages 503–507, September 2005.
- [XSA03] J. Xue, P. Stuedi, and G. Alonso. ASAP: An Adaptive QoS Protocol for Mobile Ad Hoc Networks. In *Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, volume 3, pages 2616–2620, September 2003.
- [XSLC00] H. Xiao, W.K.G. Seah, A. Lo, and K.C. Chua. A Flexible Quality of Service Model for Mobile Ad-Hoc Networks. In *Proceedings of the IEEE 51st Vehicular Technology Conference (VTC)*, volume 1, pages 445–449, May 2000.
- [ZML05] Bosheng Zhou, A. Marshall, and Tsung-Han Lee. A Cross-Layer Architecture for DiffServ in Mobile Ad-Hoc Networks. In *Proceedings of the International Conference on Wireless Networks, Communications and Mobile Computing*, volume 2, pages 833–838, June 2005.

Paper B

QoS Architecture in Ad Hoc Networks

Tor K. Moseng and Øivind Kure

Proceedings of the 2nd International Conference on Communications and Networking in China (ChinaCom)

Shanghai, China, August, 2007

Is not included due to copyright

Paper C

QoS Architecture in Ad Hoc Networks: Effects of Shadow Classes and ECN to Regulate the Load

Tor K. Moseng and Øivind Kure

Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)

Las Vegas, USA, March, 2008

Is not included due to copyright

Paper D

Evaluation of Path Dependent Scheduling in Ad Hoc Networks: a Suitable Fairness Mechanism?

Tor K. Moseng and Øivind Kure

Proceedings of the Sixth International Conference on Wireless On-demand Network Systems and Services (WONS)

Snowbird, USA, February, 2009

Is not included due to copyright

III

THESIS APPENDIX

PAPER E: QOS ARCHITECTURE IN AD HOC NETWORKS: SENSITIVITY ANALYSIS OF THE ADMISSION CONTROL

Tor K Moseng, Øivind Kure
Centre for Quantifiable Quality of Service in Communication Systems
Norwegian University of Science and Technology,
Trondheim, Norway
{torkjeti, okure}@q2s.ntnu.no

Abstract In this paper, we do a sensitivity analysis of the admission control's settings, which is a part of our proposed QoS architecture. Originally, the parameters were chosen to protect the real-time classes, but no thorough analysis was provided. Here, we reason on the settings by inspection and through simulations. We show that the highest prioritized EF class is rather insensitive to the range of settings. The AF and BE classes, on the other hand, experience large variations. However, the variation is explained through an analysis of the AF goodput.

Keywords: Ad Hoc Networks, Admission Control, QoS

1. Introduction

An ad hoc network consists of nodes that communicate without any infrastructure, which makes the nodes act as routers for each other. The shared wireless channel means that the available bandwidth will be divided between the nodes. Also, the wireless environment gives time-varying conditions that, along with mobility, cause capacity and traffic load variations. As in any network, Quality of Service (QoS) mechanisms are needed in order to differentiate traffic with different performance requirements (e.g. voice vs. e-mail). However, the variability of the wireless channel increases the complexity of QoS provisioning.

In [MK07b] we analyzed a QoS architecture based on the Differentiated Services (DiffServ) architecture [BBC⁺98]. It has an admission control based on the concept of shadow classes and monitoring of these. A new flow is sent in its designated class' shadow class for a predefined time. Based on the experienced QoS and the QoS requirements, the flow is either admitted or dropped. The designated class and its associated shadow class use the same scheduling mechanisms, but are differentiated by different drop probabilities in the buffers. Due to the dynamic topology, congestion control is needed. We included the Explicit Congestion Notification (ECN) [RFB01]

for congestion control in [MK08]. ECN controls the traffic by preempting flows. Different levels of the admission control and congestion control were analyzed to find trade-offs between utilization of the resources and QoS provisioning.

Both the admission control and the congestion control are based on the probabilistic Random Early Detection (RED) mechanism [FJ93] in the buffers, which depends on the traffic load. When the defined thresholds are reached, the packets are either dropped or marked depending on the flows' status (admitted or not). The thresholds were originally set to improve the real-time classes protection, but no thorough analysis was given. In order to see how the QoS architecture depends on these buffer parameters, we provide a sensitivity analysis around the thresholds in this paper. However, to decrease the complexity, we only focus on the EF class, and look into the effects of varying its buffer parameters.

The paper is organized as follows: Section 2 reviews some related work. In Section 3, we reason upon the parameter setting, before the sensitivity is analysed in Section 4. Finally, the paper is concluded in Section 5.

2. Related Work

There are many proposals on QoS support for ad hoc networks in the literature today. Two of these proposals, each with a different approach, are INSIGNIA [LAZC00] and SWAN [ACVS02]. INSIGNIA is an Integrated Services (IntServ) [BCS94] motivated, end-to-end resource reservation based protocol, using in-band signalling. SWAN is a class based approach that uses end-to-end probing for admission control of the real-time packets, and rate control for the best-effort packets. No per-flow information is stored along the path, which makes it more scalable. Our QoS architecture is an augmented DiffServ architecture with an admission control based on shadow classes, and congestion control using ECN with preemption [MK07b, MK08]. ECN is also used for congestion control in SWAN.

Figure 1 shows the layered QoS architecture. During the admission phase, a new flow is sent in its designated class' associated shadow class. All classes that need admission control will have an associated shadow class - in our case, the EF and AF classes. The shadow class uses the same buffer and scheduling mechanisms as its original class, but is differentiated from the original class by higher drop probabilities in the buffers through the RED mechanism. The source monitors the flow and compares the QoS requirements from the application and the QoS measured by the receiver. The flow is then either admitted to its original class, or dropped.

To avoid congestions in the network, we use, as already mentioned, ECN as congestion control in the network. Packets are marked by RED. The source monitor will after a number of ECN-marked packets preempt its flow. There will thus be two mechanisms based on the same statistical RED mechanism; new flows are controlled in shadow classes, while admitted flows are controlled through ECN. The control of the BE traffic will depend on both the traffic load in the BE class and the higher priority classes. Since it does not have any QoS requirements per se, it is only regulated by ECN in the network.

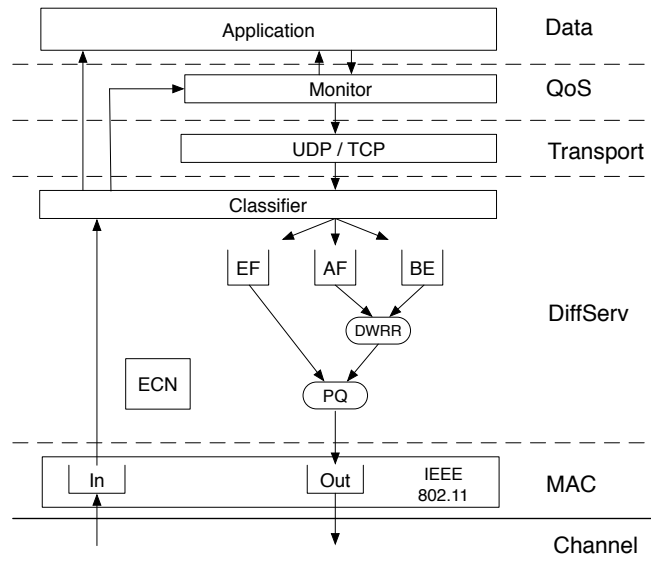


Figure 1. The QoS Architecture

3. Sensitivity Settings

The original buffer thresholds used in our proposed QoS architecture were set after some trial simulations to improve the conditions for the admitted real-time traffic. But no thorough analysis was performed. It is not the purpose of this sensitivity analysis to find the optimal parameter settings for maximizing the performance, but rather reason upon the choices and see how sensitive the traffic control mechanisms of the QoS architecture are to the buffer parameters.

A full analysis will have high complexity due to the dependencies involved; the AF class will depend on the EF class through the priority scheduler, while the control of BE traffic depends on the traffic volume in both of the two higher classes. In addition, all classes will depend on the active nodes within range, independent of priority since the IEEE 802.11 MAC [IEE07] is used. To reduce the analysis' complexity, only the EF class is considered here. By only considering the EF class, we can isolate the effects of the analysis, and avoid many interrelations in the network.

The parameters controlling the traffic consist of the buffer size and the RED thresholds. There are four RED thresholds involved: min and max for the shadow class, and min and max for the original class. The original class' thresholds are used by ECN to mark packets, while the shadow class' thresholds give the drop probability directly related to the admission of new flows. The marking of the admitted EF traffic clearly depends on the amount of traffic admitted by the EF admission controller. To further reduce the complexity and dependencies in this paper, we therefore concentrate on the shadow class' thresholds (i.e. the admission control) and the buffer size. The

Table 1. The Sensitivity Settings

	Buffer size (bytes / pck)	Min shadow (rel. size / pck)	Max shadow (rel. size / pck)
Setting 1	1024 / 17	0.25 / 4	0.75 / 12
Setting 2	1024 / 17	0.50 / 8	0.75 / 12
Setting 3	1024 / 17	0.625 / 10	0.75 / 12
Setting 4	512 / 8	0.25 / 2	0.75 / 6
Setting 5	512 / 8	0.50 / 4	0.75 / 6
Setting 6	512 / 8	0.625 / 5	0.75 / 6

min and max ECN thresholds are kept at 0.75 and 1.0 relative of the buffer size, respectively, for all settings.

We divide the analysis into two parts; in the first part we investigate the effects of changing the buffer size, while in the second part we change the thresholds. The buffer size originally used is 1 kB. A larger buffer will only give higher delays not suited for the EF class' application data. We therefore try a smaller buffer size (512 bytes) to see how this affects the class.

In [MK08] we found a high blocking of new EF flows. Only 44 % of the EF flows were admitted. To get a feeling on where to place these thresholds we simplify our situation to an M/M/1 queueing model from traditional telecommunication theory. This model does not fit the reality in our network, but serves to picture the thresholds' strictness for the traffic load targeted. Using the steady state probabilities from the M/M/1 model, we find that the probability of having more packets in the buffer than the min threshold is above 3 % for a traffic load of 0.5. Already at a 0.4 traffic load, the probability of reaching the min threshold is above 1 %. For the load targeted (0.6 - 0.7), the thresholds are therefore set too low. However, the max admission control threshold should not overlap with the min ECN threshold to give the admitted flows better protection. Therefore, we keep the max threshold, but increase the min threshold for the admission control. By trying a min threshold of 8 and 10 packets in two different settings, we can cope with a traffic load of 0.75 and 0.70, respectively, given no more than 3 % probability of exceeding this min threshold.

When we decrease the buffer size, we keep the same thresholds relative to the new buffer size. This scenario will be more strict since fewer packets are allowed before the probability of regulation becomes high, but indicates the sensitivity of the buffer size. Table 1 summarizes the parameter settings. One EF packet is 60 bytes, including RTP, UDP and IP headers.

For comparisons, setting 1 is the original setting from previous papers [MK08]. Figure 2 illustrates the table's settings by showing the shadow class' and original class' curves against the relative buffer size. Figure 2(a) corresponds to settings 1 and 4. The increased min threshold for the settings 2 and 5 is shown in Figure 2(b), while Figure 2(c) gives the parameters for the settings 3 and 6. The EF class emulates VoIP traffic over RTP and UDP; 10 bytes per sample and 2 samples per packet are used according to the G.729 audio data compression algorithm [ITU07a]. In the AF class we have CBR sources that emulate video applications with a 128 kbps data rate

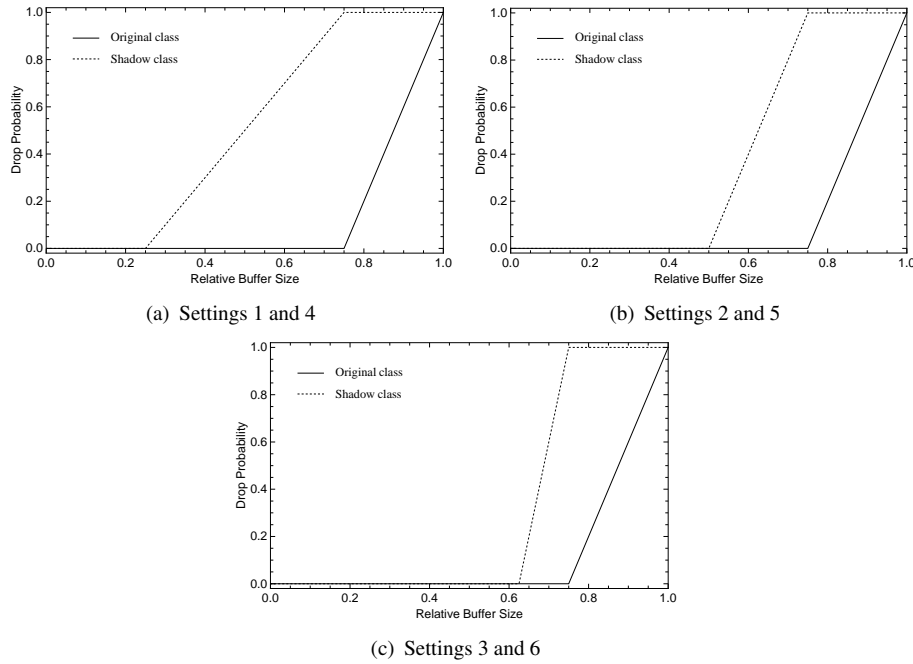


Figure 2. The buffer’s regulation probability

Table 2. Changing the buffer size

	Adm. ratio		ECN regulation		Rel. utilization		
	EF	AF	EF	AF	EF	AF	BE
Setting 1	0.44	0.27	0.39	0.43	93%	90%	100%
Setting 4	0.42	0.27	0.32	0.58	100%	100%	83%

and 256 bytes packets over RTP and UDP [ITU07b]. The admission control is set according to the QoS requirements specified by ITU-T [ITU01]. The BE class consists of TCP traffic. The maximum TCP window was set to 4 packets; higher windows gave no effect on the results. Other simulation parameters are kept the same as originally.

4. Sensitivity Analysis

In this analysis we focus on the admission ratio and utilization of the resources in order to see the performance’s sensitivity to the admission control parameters. The utilization will in this context serve as a measure of how the settings affect the amount of traffic in each class. Table 2 compares the settings 1 and 4, which gives the effect of a reduced EF buffer size. The ECN regulation ratio gives the ratio of preempted flows from the admitted flows, while the utilization is given per class and relative to the best setting in terms of number of packets delivered. A utilization of 100 % only means that this setting delivers most packets compared to the other settings in the table for that particular class.

Table 3. Changing the min threshold, buffer size = 1024

	Adm. ratio		ECN regulation		Rel. utilization		
	EF	AF	EF	AF	EF	AF	BE
Setting 1	0.44	0.27	0.39	0.43	95%	94%	73%
Setting 2	0.46	0.23	0.34	0.49	100%	86%	100%
Setting 3	0.48	0.23	0.35	0.54	100%	100%	69%

There are complex relations in this network; all classes are in a way dependent on each other. When we decrease the EF buffer size, there is less room for the EF packets, and the min drop threshold becomes active on fewer packets in the buffer. The AF class will be affected by the amount of higher prioritized EF traffic, while the BE class will have an interrelation with both the EF and AF classes; ECN considers both the EF and AF buffers when regulating the BE class. Table 2 shows that there are fewer EF flows admitted into the network. However, we see that setting 4 delivers more EF traffic than setting 1 due to less ECN regulation. The more strict admission control of shadow class packets in setting 4, gives less need for ECN regulation compared to setting 1. When the EF buffer is decreased in setting 4, it will have a higher relative occupancy, and such increase the probability of ECN marked BE packets. TCP reacts to more traffic by throttling the throughput. Since the relative share of BE traffic in the network is high, the total traffic decreases - making more room for the AF class. More analysis on the AF class is given in Section 4.1. If we look at the fairness, i.e. the performance distribution, for the EF and AF flows, there are only slight differences for a decreased EF buffer. Also, for the BE class the Coefficient of Variation (CoV) is only increased from 0.56 to 0.62 considering the BE nodes' goodput.

From Table 2, we saw the effects of a change in the buffer size. Table 3 summarizes settings 1 to 3, showing the effects of changing the min threshold with a 1024 bytes EF buffer. As we increase the min threshold, more EF flows are admitted into the network. Together with less ECN regulation, the EF utilization increases. More prioritized EF traffic give worse conditions for the AF and BE classes. In setting 2, the AF class suffers most, while in setting 3 the BE class is decreased enough to allow more traffic from the AF class. The complex relations give large variations for the AF and BE classes. In Section 4.1, we find the factors that explain the variability in the AF utilization (for settings 1 to 6). The variability is continued for settings 4 to 6 in Table 4. The EF class behaves pretty much as in settings 1 to 3 with a slightly higher admission ratio and low variability in utilization. Setting 5 stands out with more strict AF admission, but less strict ECN regulation of EF and AF. It is obvious that both the AF and BE class react to the EF traffic by admitting fewer flows and lowering the throughput, respectively. It seems that for small buffer sizes, we do not get the same effects as for larger buffers due to the small bounds between the thresholds; small traffic changes may cause large variations.

Table 4. Changing the min threshold, buffer size = 512

	Adm. ratio		ECN regulation		Rel. utilization		
	EF	AF	EF	AF	EF	AF	BE
Setting 4	0.42	0.27	0.32	0.58	100%	100%	100%
Setting 5	0.44	0.23	0.26	0.46	100%	82%	62%
Setting 6	0.45	0.26	0.32	0.56	97%	89%	97%

Table 5. Simulation results for the AF class

	Packet loss	Admission ratio	Avg admitted flow duration	Avg no of flows
Setting 1	0.10	0.27	30 sec	25
Setting 2	0.16	0.23	34 sec	25
Setting 3	0.11	0.23	41 sec	24
Setting 4	0.09	0.27	37 sec	24
Setting 5	0.11	0.23	30 sec	25
Setting 6	0.11	0.26	32 sec	24

4.1 Analysing the Variability

The results in Table 3 and 4 showed variability among the AF and BE classes. The BE class runs TCP traffic that have its own congestion avoidance mechanism. This makes the TCP traffic dependent on the network's traffic level. In addition, the BE class is regulated by ECN, which is dependent on the traffic volume in the other classes. On the other hand, the AF class runs CBR traffic and is only indirectly affected by the traffic in the other classes by increasing the channel access time. We want to explain the variability in the AF utilization, and look into the major factors deciding the AF utilization. Table 5 summarizes some results for the AF class, which we use to compute the AF utilization. In this section we try to explain the AF results, and will therefore give the utilization in terms of Mbytes delivered instead of relative numbers.

The ECN regulation is reflected in Table 5 through the average flow duration for admitted flows. We see that the packet loss is relatively equal for all settings, except for setting 2. The same is the case for number of flows in the AF class. Since these parameters are relatively constant, the main factors deciding the utilization for the AF class are the admission ratio and ECN regulation, which really is no surprise. Given the CBR (128 kbps), we find the total utilization in Table 6. Both admitted and non-admitted flows add to the total traffic since non-admitted flows are run for a predefined time before being dropped by the admission control. We see that the computed utilization gives the relative values in Tables 2- 4. E.g. Table 2 compares settings 1 and 4; setting 1 (setting 4) gives 4.03 Mbytes (4.51 Mbytes), which is 90 % of the data delivered in setting 4, and given in Table 2. Also, the total calculated utilizations show the same values as obtained directly from the simulations.

To summarize, we saw a high variability in the utilization for small changes in the min threshold. However, by looking at different AF performance metrics, we were

Table 6. The AF class' utilization in Mbytes

	Admitted flows	Non-admitted flows	Calculated utilization	Simulated utilization
Setting 1	0.41	0.07	4.03	4.08
Setting 2	0.44	0.06	3.74	3.76
Setting 3	0.56	0.07	4.31	4.35
Setting 4	0.51	0.07	4.51	4.49
Setting 5	0.41	0.07	3.69	3.69
Setting 6	0.43	0.07	3.97	3.99

able to explain the utilization observed. Also, we found that the major factors deciding the utilization were the admission ratio and ECN regulation as we could expect.

5. Conclusions

We have previously proposed a QoS architecture for ad hoc networks based on the DiffServ framework. It was shown that it can be used with good results in an ad hoc network. The aim of this paper is to give a sensitivity analysis of the admission control used in the architecture. The analysis focused on the EF class to minimize the complexity involved.

The results show that the EF class is not very sensitive to these parameters. The admission ratio was only slightly increased for less strict settings, while the total amount of EF traffic was almost equal for all settings. However, we saw the complexity involved between the different classes. The TCP traffic in the BE class has a congestion avoidance mechanism which reacts on too much traffic. Also, the BE class, in which the TCP traffic was run, is dependent on the traffic volume in both the EF and AF classes. At the same time the EF and AF classes are dependent on the general traffic level for admission. These relations give, in contrast to the stable EF performance, variability in the lower classes' utilization - especially for small buffer sizes where only a few packets are allowed before the probability of regulation becomes high. However, the variation was explained through an analysis of the AF throughput.

A more comprehensive analysis around each class' thresholds and their interrelations is for future work. Also, in relation to our QoS architecture there are other issues for future work. Two such issues are the inclusion of a QoS enabled MAC protocol, which would improve the performance by providing differentiation on the MAC as well, and investigations concerning the interconnection with fixed networks.

References

- [ACVS02] G.-S. Ahn, A.T. Campbell, A. Veres, and L.-H. Sun. SWAN: Service Differentiation in Stateless Wireless Ad Hoc Networks. In *Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, volume 2, pages 457–466, June 2002.
- [BBC⁺98] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. An Architecture for Differentiated Services. IETF RFC 2475, December 1998.

- [BCS94] R. Braden, D. Clark, and S. Shenker. Integrated Services in the Internet Architecture: an Overview. IETF RFC 1633, June 1994.
- [FJ93] S. Floyd and V. Jacobson. Random Early Detection Gateways for Congestion Avoidance. *IEEE/ACM Transactions on Networking (TON)*, 1(4):397–413, August 1993.
- [IEE07] IEEE Computer Society. IEEE Standard for Information Technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications (IEEE Std 802.11-2007). IEEE Standard, [Online] Available: <http://www.ieee802.org/11/>, March 2007.
- [ITU01] ITU-T Study Group 12 (2001-2004) under the WTSA Resolution 1 procedure. ITU-T Recommendation G.1010: End-user Multimedia QoS Categories, November 2001.
- [ITU07a] ITU-T Study Group 16 (2005-2008) under the ITU-T Recommendation A.8 procedure. ITU-T Recommendation G.729: Coding of Speech at 8 Kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP), January 2007.
- [ITU07b] ITU-T Study Group 16 (2005-2008) under the ITU-T Recommendation A.8 procedure. ITU-T Recommendation H.264: Advanced Video Coding for Generic Audiovisual Services, November 2007.
- [LAZC00] S.-B. Lee, G.-S. Ahn, X. Zhang, and A.T. Campbell. INSIGNIA: An IP-Based Quality of Service Framework for Mobile Ad Hoc Networks. *Journal of Parallel and Distributed Computing, Special Issue on Wireless and Mobile Computing and Communications*, 60(4):374–406, April 2000.
- [MK07b] T.K. Moseng and Ø. Kure. QoS Architecture in Ad Hoc Networks. In *Proceedings of the International Conference on Communications and Networking in China (ChinaCom)*, August 2007.
- [MK08] T.K. Moseng and Ø. Kure. QoS Architecture in Ad Hoc Networks: Effects of Shadow Classes and ECN to Regulate the Load. In *Proceedings of the IEEE Wireless Communications & Networking Conference 2008 (WCNC 2008)*, pages 2450–2455, March 2008.
- [RFB01] K. Ramakrishnan, S. Floyd, and D. Black. The Addition of Explicit Congestion Notification (ECN) to IP. IETF RFC 3168, September 2001.

Bibliography

- [ACVS02] G.-S. Ahn, A.T. Campbell, A. Veres, and L.-H. Sun. SWAN: Service Differentiation in Stateless Wireless Ad Hoc Networks. In *Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, volume 2, pages 457–466, June 2002.
- [APS99] M. Allman, V. Paxson, and W. Stevens. TCP Congestion Control. IETF RFC 2581, April 1999.
- [ASSC02] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. A Survey on Sensor Networks. *IEEE Communication Magazine*, 40(8):102–114, August 2002.
- [AW05] I.F. Akyildiz and X. Wang. A survey on wireless mesh networks. *IEEE Communications Magazine*, 43(9):23–30, September 2005.
- [AZA04] A. Ahmed, S. M. H. Zaidi, and N. Ahmed. Performance Evaluation of Transmission Control Protocol in Mobile Ad Hoc Networks. In *Proceedings of the International Conference on Networking and Communications (INCC)*, pages 13–18, June 2004.
- [BBC⁺98] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. An Architecture for Differentiated Services. IETF RFC 2475, December 1998.
- [BCS94] R. Braden, D. Clark, and S. Shenker. Integrated Services in the Internet Architecture: an Overview. IETF RFC 1633, June 1994.
- [BIFD01] F. Baker, C. Iturralde, F. Le Faucheur, and B. Davie. Aggregation of RSVP for IPv4 and IPv6 Reservations. IETF RFC 3175, September 2001.
- [BKS⁺00] L. Breslau, E.W. Knightly, S. Shenker, I. Stoica, and H. Zhang. Endpoint Admission Control: Architectural Issues and Performance. *ACM SIGCOMM Computer Communication Review*, 30(4):57–69, October 2000.
- [BMJ⁺98] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, and J. Jetcheva. A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols. In *Proceedings of the 4th annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, pages 85–97, October 1998.
- [BZB⁺97] R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin. Resource ReSerVation Protocol (RSVP) - Version 1 Functional Specification. IETF RFC 2205, September 1997.
- [CDL05] C. Chaudet, D. Dhoutant, and I.G. Lassous. Performance issues with IEEE 802.11 in ad hoc networking. *IEEE Communications Magazine*, 43(7):110–116, July 2005.

- [CDS74] V. Cerf, Y. Dalal, and C. Sunshine. Specification of Internet Transmission Control Program. IETF RFC 675, december 1974.
- [CJ03] T. Clausen and P. Jacquet. Optimized Link State Routing Protocol (OLSR). IETF RFC 3626, October 2003.
- [CMC99] M.S. Corson, J.P. Macker, and G.H. Cirincione. Internet-based Mobile Ad Hoc Networking. *IEEE Internet Computing*, 3(4):63–70, 1999.
- [CMM04] C.T. Calafate, P. Manzoni, and M.R. Malumbres. Assessing the Effectiveness of IEEE 802.11e in Multi-Hop Mobile Network Environments. In *Proceedings of the IEEE Computer Society's 12th Annual International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunications Systems (MASCOTS)*, pages 205–212, October 2004.
- [CP08] I. Chakeres and C. Perkins. Dynamic MANET On-demand (DYMO) Routing. IETF Internet-Draft, June 2008.
- [CSNV05] S. Crisóstomo, S. Sargento, M. Natkaniec, and N. Vicari. A QoS Architecture Integrating Mobile Ad-Hoc and Infrastructure Networks. In *Proceedings of the 3rd ACS/IEEE International Conference on Computer Systems and Applications*, pages 897–903, 2005.
- [Dav03] B. Davie. Deployment Experience with Differentiated Services. In *Proceedings of the ACM SIGCOMM workshop on Revisiting IP QoS: What have we learned, why do we care?*, pages 131–136, 2003.
- [DBEB06] M. Duke, R. Braden, W. Eddy, and E. Blanton. A Roadmap for Transmission Control Protocol (TCP) Specification Documents. IETF RFC 4614, September 2006.
- [DCB⁺02] B. Davie, A. Charny, J.C.R. Bennett, K. Benson, J.Y. Le Boudec, W. Courtney, S. Davari, V. Firoiu, and D. Stiliadis. An Expedited Forwarding PHB (Per-Hop Behavior). IETF RFC 3246, March 2002.
- [DR04a] M.C. Domingo and D. Remondo. A Cooperation Model between Ad Hoc Networks and Fixed Networks for Service Differentiation. In *Proceedings of 29th Annual IEEE International Conference on Local Computer Networks*, pages 692–693, November 2004.
- [DR04b] M.C. Domingo and D. Remondo. Analysis of VBR VoIP Traffic for Ad Hoc Connectivity with a Fixed IP Network. In *Proceedings of IEEE Vehicular Technology Conference (VTC)*, volume 4, pages 2834–2837, September 2004.
- [DR05] M.C. Domingo and D. Remondo. A Cooperation Model and Routing Protocol for QoS Support in Ad Hoc Networks Connected to Fixed IP Networks. In *Proceedings of Advanced Industrial Conference on Telecommunications/Service Assurance with Partial and Intermittent Resources Conference/E-Learning on Telecommunications Workshop (AICT/SAPIR/ELETE)*, pages 390–395, July 2005.
- [DR08] M.C. Domingo and D. Remondo. QoS Support between Ad Hoc Networks and Fixed IP Networks. *Computer Communications*, 31(11):2646–2655, July 2008.
- [EAS07] A.A. Elmangoh, M.A. Ashibani, and F.B. Shatwan. The Interworking between EDCA 802.11e and DiffServ. In *Proceedings of the IEEE International Performance, Computing, and Communications Conference (IPCCC)*, pages 545–550, April 2007.

- [EKR00] V. Elek, G. Karlsson, and R. Ronngren. Admission Control Based on End-to-End Measurements. In *Proceedings of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, volume 2, pages 623–630, March 2000.
- [Fed] Federal Communications Commission (FCC). Part 15 - Radio Frequency Devices. [Online] Available: <http://www.fcc.gov/oet/info/rules/>.
- [FJ93] S. Floyd and V. Jacobson. Random Early Detection Gateways for Congestion Avoidance. *IEEE/ACM Transactions on Networking (TON)*, 1(4):397–413, August 1993.
- [FJL00] M. Frodigh, P. Johansson, and P. Larsson. Wireless Ad Hoc Networking - The Art of Networking Without a Network. *Ericsson Review*, (4):248–263, January 2000.
- [FK06] S. Floyd and E. Kohler. Profile for Datagram Congestion Control Protocol (DCCP) Congestion Control ID 2: TCP-like Congestion Control. IETF RFC 4341, March 2006.
- [FKP06] S. Floyd, E. Kohler, and J. Padhye. Profile for Datagram Congestion Control Protocol (DCCP) Congestion Control ID 3: TCP-Friendly Rate Control (TFRC). IETF RFC 4342, March 2006.
- [Flo94] S. Floyd. TCP and Explicit Congestion Notification. *ACM Computer Communication Review*, 24(5):10–23, 1994.
- [Gei02] J. Geier. *Wireless LANs - Implementing High Performance IEEE 802.11 Networks*. Sams Publishing, second edition, 2002.
- [Gei04] J. Geier. The State of Wireless LANs. Technical report, Network World, Inc., 2004.
- [GGH00] P. Guangyu, M. Gerla, and X. Hong. LANMAR: Landmark Routing for Large Scale Wireless Ad Hoc Networks with Group Mobility. In *Proceedings of the ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, pages 11–18, 2000.
- [GK99] P. Gupta and P.R. Kumar. The Capacity of Wireless Networks. Technical report, University of Illinois, 1999.
- [GK00] P. Gupta and P.R. Kumar. The Capacity of Wireless Networks. *IEEE Transactions on Information Theory*, 46(2):388–404, March 2000.
- [GM07] A. Giovanardi and G. Mazzini. Fairness Investigation in Ad Hoc Networks. In *Proceedings of International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pages 325–328, December 2007.
- [GRG⁺06] S. Ganu, K. Ramachandran, M. Gruteser, I. Seskar, and J. Deng. Methods for Restoring MAC Layer Fairness in IEEE 802.11 Networks with Physical Layer Capture. In *Proceedings of the 2nd International Workshop on Multi-Hop Ad Hoc Networks: from Theory to Reality*, pages 7–14, 2006.
- [GTB99] M. Gerla, K. Tang, and R. Bagrodia. TCP Performance in Wireless Multi-Hop Networks. In *Proceedings of the Second IEEE Workshop on Mobile Computing Systems and Applications (WMCSA)*, pages 41–50, February 1999.
- [HBWW99] J. Heinanen, F. Baker, W. Weiss, and J. Wroclawski. Assured Forwarding PHB Group. IETF RFC 2597, June 1999.

- [HIT07] L. Hanzo-II and R. Tafazolli. A Survey of QoS Routing Solutions for Mobile Ad Hoc Networks. *IEEE Communications Surveys & Tutorials*, 9(2):50–70, 2007.
- [HK07] A. Hamidian and U. Körner. Providing QoS in Ad Hoc Networks with Distributed Resource Reservation. In *Proceedings of the 20th International Teletraffic Congress (ITC20)*, June 2007.
- [HMBT04] M.A. Haq, M. Matsumoto, J.L. Bordim, and M. Kosuga S. Tanaka. Admission Control and Simple Class Based QoS Provisioning for Mobile Ad Hoc Network. In *Proceedings of the IEEE Vehicular Technology Conference (VTC)*, volume 4, pages 2712–2718, September 2004.
- [HP01] Z.J. Haas and M.R. Pearlman. The Performance of Query Control Schemes for the Zone Routing Protocol. *IEEE/ACM Transactions on Networking*, 9(4):427–438, August 2001.
- [HS01] H.-Y. Hsieh and R. Sivakumar. Improving Fairness and Throughput in Multi-Hop Wireless Networks. In *Proceedings of the First International Conference on Networking - Part 1*, pages 569–578, 2001.
- [Hun] Hung-ying Tyan et al. J-sim network simulator. [Online] Available: <http://sites.google.com/site/jsimofficial/>.
- [IEE04] IEEE Computer Society and the IEEE Microwave Theory and Techniques Society. IEEE Standard for Local and metropolitan area networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems (IEEE Std 802.16-2004). IEEE Standard, [Online] Available: <http://www.ieee802.org/16/>, October 2004.
- [IEE07] IEEE Computer Society. IEEE Standard for Information Technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications (IEEE Std 802.11-2007). IEEE Standard, [Online] Available: <http://www.ieee802.org/11/>, March 2007.
- [IET] IETF. MANET Working Group Charter. [Online] Available: <http://www.ietf.org/html.charters/manet-charter.html>.
- [Inf] Information Society Technologies Programme (IST). The TEQUILA project. [Online] Available: <http://www.ist-tequila.org/>.
- [Inf81] Information Sciences Institute at USC, for DARPA. Transmission Control Protocol. IETF RFC 793, September 1981.
- [ITU94] ITU-T Study Group 2 (1993-1996) under the WTSA Resolution 1 procedure. ITU-T Recommendation E.800: Terms and definitions related to quality of service and network performance including dependability, August 1994.
- [ITU01] ITU-T Study Group 12 (2001-2004) under the WTSA Resolution 1 procedure. ITU-T Recommendation G.1010: End-user Multimedia QoS Categories, November 2001.
- [ITU07a] ITU-T Study Group 16 (2005-2008) under the ITU-T Recommendation A.8 procedure. ITU-T Recommendation G.729: Coding of Speech at 8 Kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP), January 2007.
- [ITU07b] ITU-T Study Group 16 (2005-2008) under the ITU-T Recommendation A.8 procedure. ITU-T Recommendation H.264: Advanced Video Coding for Generic Audiovisual Services, November 2007.

- [Jac88] V. Jacobson. Congestion Avoidance and Control. *ACM SIGCOMM Computer Communication Review*, 18(4):314–329, August 1988.
- [Jac90] V. Jacobson. Modified TCP Congestion Avoidance Algorithm. Technical Report. Email to the end2end-interest mailing list. [Online] Available: <ftp://ftp.isi.edu/end2end/end2end-interest-1990.mail>, April 1990.
- [JCH84] R. Jain, D.M. Chiu, and W. Hawe. A Quantitative Measure of Fairness and Discrimination of Resource Allocation in Shared Systems. Technical report, Digital Equipment Corporation, 1984.
- [JL05] L.B. Jiang and S.C. Liew. Proportional Fairness in Wireless LANs and Ad Hoc Networks. In *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, volume 3, pages 1551–1556, March 2005.
- [JM96] D.B. Johnson and D.A. Maltz. *Mobile Computing*, chapter 5: Dynamic Source Routing in Ad Hoc Wireless Networks, pages 153–181. Kluwer Academic Publishers, 1996.
- [Joh94] D.B. Johnson. Routing in Ad Hoc Networks of Mobile Hosts. In *Proceedings of the Workshop on Mobile Computing Systems and Applications*, pages 158–163, December 1994.
- [JS03] J. Jun and M.L. Sichitiu. Fairness and QoS in Multihop Wireless Networks. In *Proceedings of Vehicular Technology Conference (VTC)*, volume 5, pages 2936–2940, October 2003.
- [JT87] J. Jubin and J.D. Tornow. The DARPA Packet Radio Network Protocols. *Proceedings of the IEEE*, 75(1):21–32, January 1987.
- [Kah77] R. Kahn. The Organization of Computer Resources into a Packet Radio Network. *IEEE Transactions on Communications*, 25(1):169–178, January 1977.
- [KHF06] E. Kohler, M. Handley, and S. Floyd. Datagram Congestion Control Protocol (DCCP). IETF RFC 4340, March 2006.
- [KKZ00] F.P. Kelly, P.B. Key, and S. Zachary. Distributed Admission Control. *IEEE Journal on Selected Areas in Communications*, 18(12):2617–2628, December 2000.
- [KS78] L. Kleinrock and J.A. Silvester. Optimum Transmission Radii for Packet Radio Networks or Why Six is a Magic Number. In *Proceedings of the IEEE National Telecommunications Conference*, December 1978.
- [Kuo74] F.F. Kuo. The ALOHA System. *ACM SIGCOMM Computer Communication Review*, 4(1):5–8, January 1974.
- [Kuz05] A. Kuzmanovic. The Power of Explicit Congestion Notification. In *Proceedings of the ACM SIGCOMM*, August 2005.
- [LAC01] S.-B. Lee, G.-S. Ahn, and A.T. Campbell. Improving UDP and TCP performance in mobile ad hoc networks with INSIGNIA. *IEEE Communications Magazine*, 39(6):156–165, June 2001.
- [LAZC00] S.-B. Lee, G.-S. Ahn, X. Zhang, and A.T. Campbell. INSIGNIA: An IP-Based Quality of Service Framework for Mobile Ad Hoc Networks. *Journal of Parallel and Distributed Computing, Special Issue on Wireless and Mobile Computing and Communications*, 60(4):374–406, April 2000.

- [LBC⁺01] J. Li, C. Blake, D.S.J. De Couto, H.I. Lee, and R. Morris. Capacity of Ad Hoc Wireless Networks. In *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking (MobiCom)*, pages 61–69, 2001.
- [LC05] X. Li and L. Cuthbert. Multipath QoS Routing of supporting DiffServ in Mobile Ad hoc Networks. In *Proceedings of the ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD/SAWN)*, pages 308–313, May 2005.
- [LYZW06] W. Liu, J. Yi, D. Zhao, and J. T. Wen. A Smooth Control to Avoid Congestion for Wireless Network. In *Proceedings of the IEEE International Conference on Networking, Sensing and Control (ICNSC)*, pages 613–618, April 2006.
- [MAF05] A. Medina, M. Allman, and S. Floyd. Measuring the Evolution of Transport Protocols in the Internet. *ACM SIGCOMM Computer Communication Review*, 35(2):37–52, April 2005.
- [Mas04] K. Mase. Toward Scalable Admission Control for VoIP Networks. *IEEE Communications Magazine*, 2004.
- [MCG⁺03] E. Mykoniati, C. Charalampous, P. Georgatsos, T. Damlatis, D. Goderis, P. Trimintzios, G. Pavlou, and D. Griffin. Admission Control for Providing QoS in DiffServ IP Networks: The TEQUILA Approach. *IEEE Communications Magazine*, 41(1):38–44, January 2003.
- [MD01] M.K. Marina and S.R. Das. On-demand multipath distance vector routing for ad hoc networks. In *Proceedings of the International Conference on Network Protocols*, pages 14–23, November 2001.
- [MHT07] L. Mamatas, T. Harks, and V. Tsaoussidis. Approaches to Congestion Control in Packet Networks. *Journal of Internet Engineering*, 1(1), January 2007.
- [MK03] Y.L. Morgan and T. Kunz. PYLON: An Architectural Framework for Ad-Hoc QoS Interconnectivity with Access Domains. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences (HICSS)*, January 2003.
- [MK07a] T.K. Moseng and Ø. Kure. DiffServ in Ad Hoc Networks. *Wireless Systems and Mobility in Next Generation Internet, Lecture Notes in Computer Science (LNCS)*, 4396/2007:113–125, February 2007.
- [MK07b] T.K. Moseng and Ø. Kure. QoS Architecture in Ad Hoc Networks. In *Proceedings of the International Conference on Communications and Networking in China (ChinaCom)*, August 2007.
- [MK08] T.K. Moseng and Ø. Kure. QoS Architecture in Ad Hoc Networks: Effects of Shadow Classes and ECN to Regulate the Load. In *Proceedings of the IEEE Wireless Communications & Networking Conference 2008 (WCNC 2008)*, pages 2450–2455, March 2008.
- [MWH01] M. Mauve, A. Widmer, and H. Hartenstein. A Survey on Position-based Routing in Mobile Ad Hoc Networks. *IEEE Network*, 15(6):30–39, 2001.
- [PB94] C.E. Perkins and P. Bhagwat. Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers. *ACM SIGCOMM Computer Communication Review*, 24(4):234–244, October 1994.

- [PBR01] C.E. Perkins and E.M. Belding-Royer. Quality of Service for Ad Hoc On-Demand Distance Vector Routing. IETF Internet-Draft, November 2001.
- [PBRD03] C. Perkins, E. Belding-Royer, and S. Das. Ad Hoc On-Demand Distance Vector (AODV) Routing. IETF RFC 3561, July 2003.
- [PFTK98] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose. Modeling TCP Throughput: a Simple Model and its Empirical Validation. Technical report, UMASS CMPSCI Tech Report TR98-008, February 1998.
- [Pos80] J. Postel. User Datagram Protocol. IETF RFC 768, August 1980.
- [PR99] C.E. Perkins and E.M. Royer. Ad-Hoc On-Demand Distance Vector Routing. In *Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications (WMCSA)*, pages 90–100, February 1999.
- [PRDM01] C.E. Perkins, E.M. Royer, S.R. Das, and M.K. Marina. Performance Comparison of Two On-demand Routing Protocols for Ad Hoc Networks. *IEEE Personal Communications*, 8(1):16–28, February 2001.
- [RFB01] K. Ramakrishnan, S. Floyd, and D. Black. The Addition of Explicit Congestion Notification (ECN) to IP. IETF RFC 3168, September 2001.
- [RG03] Z. Ren and W. Guo. Unicast Routing in Mobile Ad Hoc Networks: Present and Future Directions. In *Proceedings of the Fourth International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)*, pages 340–344, August 2003.
- [rGPPG04] 3rd Generation Partnership Project (3GPP). Quality of Service (QoS) Concept and Architecture (Release 6). Technical Specification Group Services and System Aspects, TS23.107 v6.1.0, March 2004.
- [RJ90] K.K. Ramakrishnan and R. Jain. A Binary Feedback Scheme for Congestion Avoidance in Computer Networks. *ACM Transactions on Computer Systems (TOCS)*, 8(2):158–181, May 1990.
- [RT99] E.M. Royer and C.-K. Toh. A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks. *IEEE Personal Communications*, 6(2):46–55, April 1999.
- [SJ04] K.N. Sridhar and L. Jacob. Interplay of Service Differentiation and Routing Mechanisms in MANETs: A Performance Study. In *Proceedings of the IEEE International Conference on Performance, Computing and Communications*, pages 147–151, 2004.
- [SK05] C. ShanShan and A.J. Kassler. Extending SWAN to Provide QoS for MANETs Connected to the Internet. In *Proceedings of International Symposium on Wireless Communication Systems (ISWCS)*, pages 503–507, September 2005.
- [SPG97] S. Shenker, C. Patridge, and R. Guerin. Specification of the Guaranteed Quality of Service. IETF RFC 2212, September 1997.
- [SSB99] R. Sivakumar, P. Sinha, and V. Bharghavan. CEDAR: A Core-Extraction Distributed Ad Hoc Routing Algorithm. *IEEE Journal on Selected Areas in Communications*, 17(8):1454–1465, August 1999.
- [Sun] Sun Microsystems Inc. The Java Technology Platform. [Online] Available: <http://java.sun.com/>.

- [Thea] The Contiki project. A Memory-Efficient Operating System for Embedded Smart Objects. [Online] Available: <http://www.sics.se/contiki/>.
- [Theb] The U.S. Department of Defense. Global Positioning System. [Online] Available: <http://www.gps.gov/>.
- [Tob87] F.A. Tobagi. Modeling and Performance Analysis of Multihop Packet Radio Networks. *Proceedings of the IEEE*, 75(1), January 1987.
- [Wro97] J. Wroclawski. Specification of the Controlled-Load Network Element Service. IETF RFC 2211, September 1997.
- [WSGLA08] Z. Wang, H.R. Sadjadpour, and J.J. Garcia-Luna-Aceves. A Unifying Perspective on the Capacity of Wireless Ad Hoc Networks. In *Proceedings of the 27th IEEE International Conference on Computer Communications (INFOCOM)*, pages 211–215, April 2008.
- [XSA03] J. Xue, P. Stuedi, and G. Alonso. ASAP: An Adaptive QoS Protocol for Mobile Ad Hoc Networks. In *Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, volume 3, pages 2616–2620, September 2003.
- [XSLC00] H. Xiao, W.K.G. Seah, A. Lo, and K.C. Chua. A Flexible Quality of Service Model for Mobile Ad-Hoc Networks. In *Proceedings of the IEEE 51st Vehicular Technology Conference (VTC)*, volume 1, pages 445–449, May 2000.
- [XTB⁺03] K. Xu, K. Tang, R. Bagrodia, M. Gerla, and M. Bereschinsky. Adaptive Bandwidth Management and QoS Provisioning in Large Scale Ad Hoc Networks. In *Proceedings of the IEEE Military Communications Conference (MILCOM)*, volume 2, pages 1018–1023, October 2003.
- [YK05] Y. Yang and R. Kravets. Contention-Aware Admission Control for Ad Hoc Networks. *IEEE Transactions on Mobile Computing*, 4(4):363–377, 2005.
- [ZdV05] A. Zemlianov and G. de Veciana. Capacity of Ad Hoc Wireless Networks with Infrastructure Support. *IEEE Journal on Selected Areas in Communications*, 23(3):657–667, March 2005.
- [ZML05] Bosheng Zhou, A. Marshall, and Tsung-Han Lee. A Cross-Layer Architecture for DiffServ in Mobile Ad-Hoc Networks. In *Proceedings of the International Conference on Wireless Networks, Communications and Mobile Computing*, volume 2, pages 833–838, June 2005.