Journal of Sustainable Development in Africa (Volume 10, No.2, 2008)

ISSN: 1520-5509

Clarion University of Pennsylvania, Clarion, Pennsylvania

QUALITY OF SERVICE IN MOBILE AD-HOC NETWORKS

W. Munyoka, C. Gombiro, G. Chengetanai, N. Mlambo, C. Zano, F. Mauchi
Bindura University of Science Education

Abstract

Quality of Service (QoS) is essential in Mobile Ad Hoc Networks (MANETs) in order to satisfy communication constraints, such as delay, bandwidth, probability of packet loss, delay variance (jitter). Traffic communication from such applications is treated as high-priority, i.e. as in real-time applications, which are non-delay tolerant.

This paper presents a strategy for providing flow-based QoS on top of Media Access Control (MAC) Layer differentiation mechanisms in 802.11 ad-hoc wireless networks, with respect to the number of high-priority-flows under three different circumstances.

The targeted audiences are the Southern Africa Telecommunications designers and/or system users of mobile networks such as Internet Service providers, Mobile phone providers such as Econet and Netone, Mobile GIS networks used by meteorological department of Zimbabwe, Zimbabwe Defense Forces and Mines.
We used Network Simulator (NS-2) to provide substantial support for our simulation of routing and multicast protocols over wired and wireless networks. Our simulation results reveal that differentiation service mechanisms at the 802.11 MAC layer managed to provide quality of service under low network load only; but as the traffic increases, performance of these mechanisms perform poorly.

Introduction

An ad-hoc network is an autonomous wireless network that can be formed without the need of any infrastructure or centralized administration. It is composed of stations that communicate with each other through single-hop or multi-hop paths in a peer-to-peer fashion. We are living in an increasingly wireless world and in Zimbabwe we are also increasingly getting mobile. Networks in Zimbabwe are becoming highly congested requiring the need of quality of service provision. This means that not only communications, but also the daily tasks we perform need to keep up with our mobile lifestyle.

With its increase in use in multimedia applications, there is a need for wireless network to provide some QoS guarantees for highly sensitive applications. These QoS guarantees can mainly be achieved through resource reservation and admission control.

Our focus in this paper is based on the flow reservation and admission control for quality of service (QoS) on IEEE 802.11 ad-hoc networks. The main aim being to provide per-flow QoS through reservation along the path that connects two mobile nodes. Hence, provide better QoS guarantees for real-time applications on top of MAC IEEE 802.11 differentiation services through its evaluation as an aspect of QoS in ad-hoc networks. Providing service differentiation in IEEE 802.11 ad-hoc networks has been of interest in recent years. It's one of the aspects of providing QoS in wireless networks by mainly manipulating MAC layer parameters. Most of these service differentiation mechanisms perform poorly as traffic increases in the network.

The concept of differentiation mechanisms is to minimize packet transmission collision either by varying the contention window size for different traffic categories, or by employing appropriate back-off interval calculation methods to control the channel access of wireless stations. Under light or medium load, these mechanisms are efficient in the sense that all flows should backoff for a different period of time after collision and thus the likelihood of colliding again is rare. However, when the traffic load increases, without flow admission control, collision rate cannot be significantly reduced by only adjusting backoff intervals, i.e., collision rate increases with the increase of the traffic load. This increasing collision rate is very harmful for system performance and may yield low throughput and large packet delays.

Therefore, in order to avoid the severe performance degradation under high traffic load, QoS flow reservation and admission control become critical.

Mobile Ad-hoc Network and Quality of Services

This provides a background on ad-hoc networks and review existing approaches in the provisioning of QoS in wireless ad-hoc networks. Existing approaches can be categorized into the following groups: MAC QoS, QoS-aware routing, inter-layer QoS model and QoS-aware application. We also look at admission control and flow reservation in ad-hoc networks.

Mobile Ad-hoc Networks

A mobile ad-hoc network (MANET) is a wireless network temporarily and spontaneously created by mobile stations without requiring any infrastructure or central control. Network managements and communications are typically performed in a distributed manner. Some unique features make ad-hoc networks distinct from other types of wireless networks such as wireless LANs.

The first peculiarity is infrastructure-less, i.e. there is no pre-existing hardware like base stations in traditional cellular networks or any centralized mechanism managing the network. Ad-hoc networks are usually deployed in emergent and temporary situations such as accidents or public gatherings, where mobile stations may join the network at will, move around, or become disconnected at any time. Global synchronization is hard to achieve in such situations. And it is unrealistic to expect such a network to be fully connected, in which case a mobile station can communicate directly with every other nodes in the network via wireless channels (see Figure 1). As a result, the second important feature emerges - multihop communication. Each node in the network has to take the responsibility of relaying packets for its peers and a packet may traverse multiple nodes before it reach the destination.

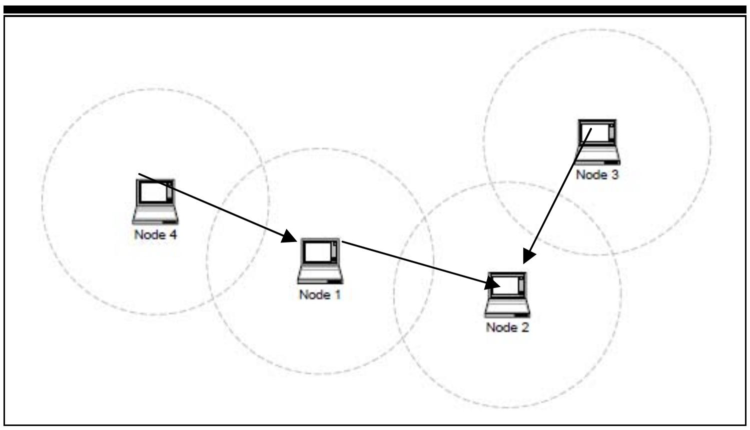

Figure 1: Example of a MANET

Quality of Service (QoS)

Quality of Service (QoS) refers to a set of service requirements that needs to be met by the network while transporting a packet stream from a source to its destination [17]. Informally, it refers to the probability of a packet passing between two points in the network. The network is expected to guarantee a set of measurable pre-specified service attributes to the users in terms of end-to-end performance, such as delay, bandwidth, probability of packet loss, delay variance (jitter), power consumption etc.

QoS metrics

QoS metrics are base parameters of quality for a network. QoS parameters include bandwidth, delay, jitter, security, network availability, and battery life and packet loss.

The important QoS metrics for multimedia applications are delay, jitter, loss, and throughput. End-to-end delay is the time between the arrival of a packet and its successful delivery to the receiver. Another metric, access delay, is the time between packet arrival and packet transmission by the sender. Jitter is the variation of delay and is an important metric for multimedia applications. Bandwidth is the measure of data transmission capacity and influences throughput, which is the amount of data successfully transmitted and received in unit time. Some of the data is lost in transit, and reducing the loss rate is an important QoS goal as well.

Media Access Control-QoS

The MAC approach provides QoS support at the media access control (MAC) layer. Radio channels are shared media, and can be shared differently to provide service differentiation for instance by assigning larger slots for higher priority packets. The 802.11 MAC protocol parameters, such as the Interframe spacing (IFS), Contention Window (CW), and Backoff Integer (BI) have been suggested for QoS support.

Best-effort distributed MAC controllers are widely used in wireless ad-hoc networks. The IEEE 802.11 Distributed Coordination Function (DCF) is a good example of a best-effort distributed MAC. The Enhanced Distributed Coordination Function (EDCF) is a growing IEEE 802.11 alternative that facilitates prioritized packet transmission [22]. Recently, there have been a number of proposals to support service differentiation at the MAC layer using distributed control schemes. [2]

Inter-layer QoS model Solutions

The inter-layer QoS model operates over different routing mechanisms and various media access layer. The inter-layer QoS model approach follows the flavor of QoS solutions for fixed topology networks namely by viewing routing mechanisms as one distinct component that can interact with the QoS model. This approach has started by importing solutions from fixed wired topology networks as in Flexible QoS Model for mobile ad-hoc networks (FQMM) [26].FQMM combines a reservation mechanism for high-priority traffic (IntServ) and a service differentiation (DiffServ) for low-priority data. Other approaches have realized the unique characteristics of ad-hoc networks as in in-band Signaling (INSIGNIA) [15] and Stateless Wireless Ad-hoc networks (SWAN) [3]. INSIGNIA uses an in-band signaling protocol for distribution of QoS information. The information is included in the IP headers of the data packets, and the available resources are calculated at each station the packet traverses so that a QoS decision can be made.

Types of QoS Applications

QoS applications can be classified into Real-time (RT) and Elastic applications. RT- applications need packets to arrive within certain time limits, and will disregard packets arriving past that time. Elastic applications can tolerate delays of arrival, and can afford to wait for packets.

QoS Challenges in MANETs

The dynamics of ad-hoc networks in terms of node mobility, limited battery power, and variable radio quality, make it difficult for real-time applications with appropriate QoS. The network dynamics also make it difficult to assign a central controller to maintain connection state and reservations. Major QoS challenges facing ad-hoc networks can be summarized as follows:

QoS challenges due to mobility of nodes. These challenges make it difficult to maintain resources on specific routes. The network dynamics impose inherent limitations to QoS promises in terms of connectivity, and robustness.

QoS challenges due to unpredictable link properties such as interference with other wireless devices, signal fading, or hidden node issues. This problem results in variant resources even on a fixed route and even assuming no mobility, for instance, due to interference with, potentially, wireless devices outside the ad-hoc network. The unpredictability of a wireless link causes potential variations in the link capacity, and therefore, inherent limitations on the expected QoS guaranties. This challenge makes flow reservation difficult to attain in ad-hoc networks.

QoS challenges due to limited capabilities of mobile nodes in terms of processing power, storage capacity, or energy. The limited capabilities challenge, influence, and shape the QoS design for instance by forcing a distributed approach, avoiding lookup tables, accommodating dormant devices, or adopting simpler lightweight algorithms.

QoS challenges due to the lack of central authority that can maintain central information on flows, routes, or connections. The challenge here is to design a decentralized QoS schemes.

QoS challenges due to Hidden and Exposed Terminal Problems: In a MAC layer with the traditional carrier sense multiple access (CSMA) protocol, multihop packet relaying introduces the “hidden terminal” and “exposed terminal” problems. The hidden terminal problem happens when signals of two nodes, say A and C, that are out of each other’s transmission ranges collide at a common receiver, say node B (see Figure 2). An exposed terminal problem will result from a scenario where node B attempts to transmit data to node A; while node C is transmitting to node D. In such a case, node B is exposed to the transmission range of node C and thus defers its transmission even though it would not interfere with the reception at node D (see Figure 2).

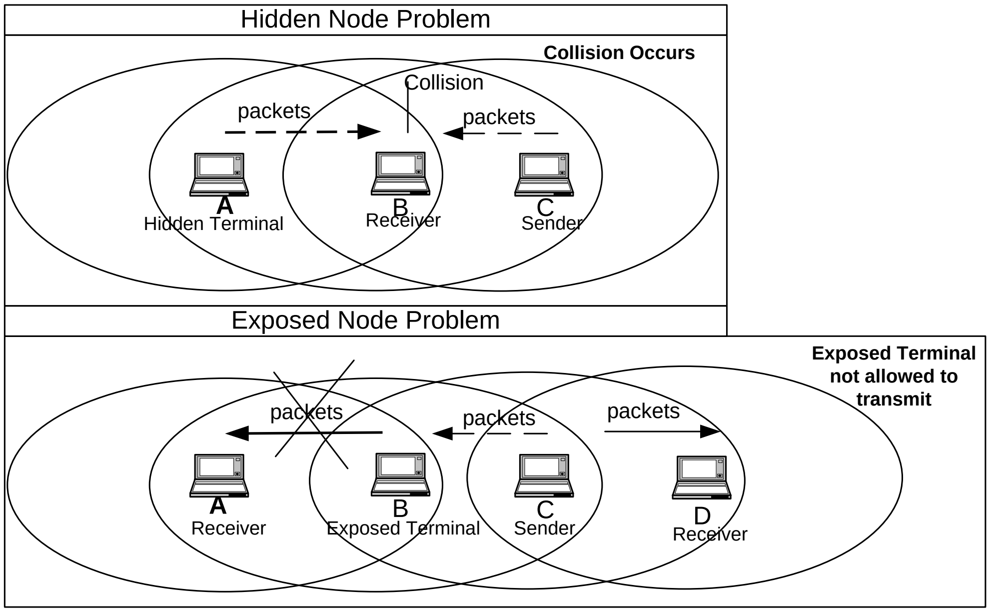

Figure 2. Illustration of Hidden and Exposed Node Problems

All these challenges lead to serious concern in the provision of quality of service in ad-hoc networks. Some of these challenges influence greatly the issue of flow reservation in ad-hoc networks.

Institute of Electrical and Electronics Engineers (IEEE) 802.11

In general, the IEEE 802.11 [23] standard covers the MAC sub-layer and the physical (PHY) layer of the OSI (Open Systems Interconnection) network reference model. Logical Link Control (LLC) sub-layer is specified in the IEEE 802.2 standard. This architecture provides a transparent interface to the higher layer users: stations may move, roam through an 802.11 wireless network and still appear as stationary to 802.2 LLC sub-layer and above. This allows existing network protocols (such as TCP/IP) to run over IEEE 802.11 wireless without any special considerations, just like if IEEE 802.3 wired Ethernet was deployed.

At PHY layer, first the IEEE provides three kinds of options in the 2.4 GHz band. The three PHY layers are an Infrared (IR) baseband PHY, a Frequency Hopping Spread Spectrum (FHSS) radio and a Direct Sequence Spread Spectrum (DSSS) radio. All three PHY layers support both 1 and 2Mbps operation. In 1999, the IEEE defined up to 11Mbps 802.11b in the 2.4 GHz free ISM (Industrial, Science, and Medical) band and up to 54Mbps 802.11a OFDM in 5GHz frequency. Ongoing 802.11g will extend 2.4GHz 802.11b PHY layer to support at least 20Mbps rate. Moreover, 802.11h will enhance 802.11a in the 5GHz band, adding indoor and outdoor channel selection for 5GHz license exempt bands in Europe. At MAC layer, ongoing 802.11e covers QoS support to the 802.11 wireless networks. 802.11i will enhance security and authentication mechanisms for 802.11 MAC.

The IEEE 802.11 MAC sub-layer defines two relative medium access coordination functions, the Distributed Coordination Function (DCF) and the optional Point Coordination Function (PCF). The transmission medium can operate both in contention mode (DCF) and contention-free mode (PCF). The IEEE 802.11 MAC protocol provides two types of transmission: asynchronous and synchronous.

The asynchronous type of transmission is provided by DCF, which implements the basic access method of the 802.11 MAC protocol. DCF is based on the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol, and should be implemented in all the stations. The synchronous service (also called contention free service) is provided by PCF, which basically implements a polling-based access method.

Distributed Coordination Function (DCF)

The basic scheme for DCF is Carrier Sense Multiple Access (CSMA). This protocol has two variants: Collision Detection (CSMA/CD) and Collision Avoidance (CSMA/CA).

A collision can be caused by two or more stations using the same channel at the same time after waiting a channel idle period, or (in wireless networks) by two or more hidden terminals emitting at the same time.

CSMA/CD is used in Ethernet (IEEE 802.3) wired networks. Whenever a node detects that the transmitted signal is different from the one on the channel, it aborts transmission, saving useless collision time. This mechanism is not possible in wireless communications, as nodes cannot listen to the channel while transmitting, due to the big difference between transmitted and received power levels. In this case, after each frame transmission the sender waits for an acknowledgment (ACK) from the receiver, as shown in Figure 3.

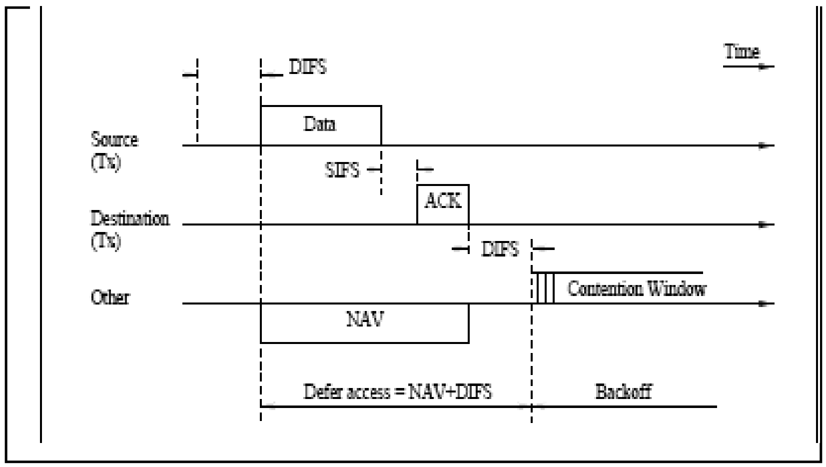

Figure 3. Basic Access Scheme

Source axis shows data transmitted by the source. The destination responds by an ACK, represented on the *Destination* axis. The third axis represents the network state, as seen by *other* nodes. Note that transmission delays are not shown. The Interframe Spacing's DIFS and SIFS will be explained later in this Section.

If no ACK was returned, a collision must have occurred and the frame is retransmitted. But this technique may waste a lot of time in case of long frames, keeping transmission going on while congestion is taking place (caused by a hidden terminal for example). This can be solved by introducing an optional RTS/CTS scheme (Request to Send and Clear to Send respectively), in addition to the previous basic scheme.

In the *optional RTS/CTS scheme*, a station sends an RTS before each frame transmission for channel reservation. The destination responds with a CTS if it is ready to receive and the channel is idle for the packet duration. When the source receives the CTS, it starts transmitting its frame, being sure that the channel is "reserved" for the frame duration. All other nodes update their *Network Allocation Vector* (NAV) at each hearing of RTS, CTS and the data frames. NAV is used for *virtual carrier sensing*, detailed in the next paragraph.

This scheme is shown in Figure 4. The overhead caused by the transmission of RTS/CTS frames becomes considerable when data frame sizes are small and sub-optimal channel usage takes place.

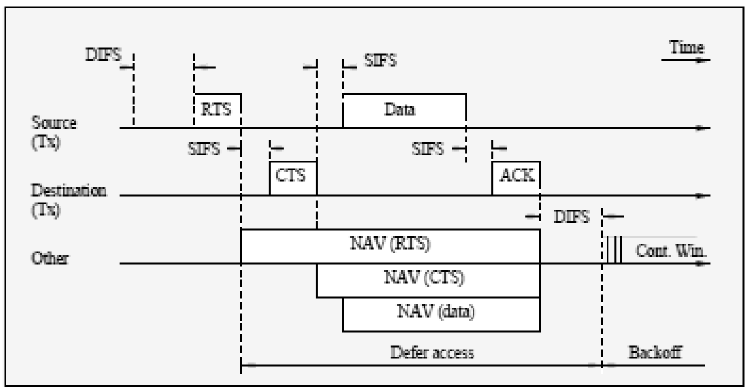

Figure 4. RTS/CTS Access Scheme

Not all packet types have the same priority. For example, ACK packets should have priority over RTS or data ones. This is done by affecting to each packet type a certain *Interframe Spacing* (IFS) before which a packet cannot be transmitted, once the channel becomes idle. In DCF two IFSs are used: Short IFS (SIFS) and DCF IFS (DIFS), where SIFS is shorter than DIFS (See Figure 3 and Figure 4). As a result, if an ACK (affected with SIFS) and a new data packet (affected with DIFS) are waiting simultaneously for the channel to become idle, the ACK will be transmitted before the new data packet (the first has to wait SIFS whereas the data has to wait DIFS.)

Carrier sensing can be performed on both layers. On the physical layer *physical carrier sensing* is done by detecting any channel activity caused by other sources. On the MAC sub-layer, *virtual carrier sensing* can be done by updating a local NAV with the value of other terminal's transmission duration. This duration is declared in data frames, RTS and CTS frames. Using the NAV, a node MAC knows when the current transmission will end. NAV is updated upon hearing an RTS from the sender and/or a CTS from the receiver, so the hidden node problem is avoided. In our work will use this mechanism to estimate the available bandwidth to be reserved.

The collision avoidance part of CSMA/CA consists of avoiding packet transmission right after the channel is sensed idle (+ DIFS time), so it won't collide with other "waiting" packets. Instead, a node with a packet ready to be transmitted waits a random time after the channel being idle for DIFS, backoff time, shown in Figure 3 and Figure 4. Backoff time of each node is decreased as long as the channel is sensed idle (during the called *contention window*). When the channel is busy, backoff time is frozen. When backoff time reaches zero, the node transmits its frame, but if the channel is sensed busy because of another "waiting" frame, the node computes a new random backoff time, with a new range. This range increases exponentially as 2^{2+i} where i (initially equal to 1) is the transmission attempt number. Therefore, the backoff time equation is:

$$\mathit{Backoff\ time} = [\, 2^{2+i} * \mathrm{rand}\ ()\,] * \mathit{Slot\_Time}$$

Where Slot time is function of some physical layer parameters, and rand () is a random function with a uniform distribution in [0, CW]. There is a higher limit for retransmission attempts i, above which the frame will be dropped. Collision avoidance is applied on data packets in the basic scheme, and on RTS packets in the RTS/CTS scheme.

All nodes have equal probability to access the channel, thus share it equally; but this method has no guarantees for queuing delays and has no service differentiation.

Service Differentiation

Service differentiation is an important aspect of providing QoS in wireless networks. In many ad-hoc network applications, such as disaster rescue, communication terminals may have different priority ranks. Many applications that are deployable in ad hoc networks, such as multimedia applications, may have different delivery requirements, i.e., low delay and jitter, and high throughput. For instance, a typical Voice over IP (VoIP) traffic session has the requirement of very low transmission delay. While multimedia-streaming traffic is more tolerant to latency than VoIP traffic, it requires more bandwidth. We can therefore label different traffic classes with different priority levels and provide service differentiation among traffic flows.

The essential problem of providing QoS in multi-hop ad-hoc networks is trying to admit as many traffic flows as possible in order to achieve high efficiency of the channel usage, while at the same time providing service quality guarantees according to traffic priority.

Service differentiation among stations in the 802.11 standard is archived by assigning different priorities in the wireless medium access to stations that contend for it. These proposals suggest modifications to the DCF mode.

These techniques can be classified according to the parameter used to archive differentiation: *DIFS*, *backoff*, *frame size*, and *RTS/CTS threshold*.

The *DIFS-based scheme* consists of configuring wireless stations with different values for this parameter according to the priority that one wishes to assign to each station. The larger the DIFS in the number of slots, the smaller the station priority. To avoid contention among stations with different priorities, the maximum contention window of a station with priority j added to $DIFS_j$ is chosen in such a way that it is never larger than $DIFS_{j+1}$ (lower priority). This guarantees that a higher priority station has no frames to send when a lower priority station starts transmitting.

The *backoff-based scheme* consists of assigning different intervals (min and max) for contention window of each station or determining how the contention window evolves along with station/flow priority, number of retransmission retrials, and other factors. In [2], the contention window intervals are calculated according to the priority established for each station. Aad et al [1] also present a mechanism that assigns different priorities for different destinations, i.e., per-flow differentiation. In [10][11], the authors propose a scheme where the priority of the next frame to be sent is included in RTS and CTS control frames, data frame, and the corresponding ACK. Since all stations in the same coverage area, hear this information can maintain a table with the current head-of-line frames of all stations that contend for the medium. The contention window interval is calculated then by each station according to the position (rank), in terms of priority, of its frame in that table. This scheme does not provide an admission control mechanism, resulting in performance degradation as the traffic load increases.

Bensaou et al [24] propose a scheme of differentiated backoff according to the estimate of its bandwidth share and the share obtained by the other stations. The main idea is to allow all stations to transmit using the default configuration if the total load is smaller than the link capacity. In case of exceeding the link capacity, each station should obtain an access proportional to sharing index previously established in the admission control.

The two schemes described below establish a coarser differentiation. In the technique based on the frame size, stations with higher priority use larger frame sizes in their transmissions. This scheme controls the time a station retains the medium after winning a contention for it.

The technique based on the *RTS/CTS threshold* consists of the use of medium reservation through the RTS/CTS handshake. Stations with threshold values larger than frame sizes of a certain flow will not use RTS/CTS. These frames will have higher collision probability and consequently a lower priority.

In our work we will mainly concentrate and evaluate DIFS-based and backoff-based schemes because they are schemes where the currently effort to provide QoS in 802.11e is based.

Admission Control

Admission control aims to provide a path, from source to destination, containing enough free resources to carry a flow, without interfering with nearby ongoing traffic.

Since we are assuming a shared medium, the Ad-hoc On-demand Distance Vector routing protocol (AODV) must be able to access bandwidth related information of every node on the path, as well as their first hop neighbors.

Admission control schemes can be broadly classified into measurement-based and calculation-based methods. In measurement-based schemes, admission control decisions are made based on the measurements of existing network status, such as throughput and delay. On the other hand, calculation-based schemes construct certain performance metrics or criteria for evaluating the status of the network. In our approach the admission is performed at the network layer of the Open Systems Interconnection (OSI).

Wireless networks generally have limited resources in terms of both device capabilities and available network bandwidth. Consequently, it is beneficial to have call admission to prevent unprovisioned traffic from being injected into the network beyond the saturation point. If a flow has rigid QoS requirements, an admission mechanism will prevent the waste of resources of both the source node itself and the whole network, if the network cannot support the flow. Furthermore, wireless communication channels are shared by all nodes within transmission range; consequently, all nodes within a transmission area contend for the limited channel bandwidth.

In a multi-hop scenario, an admitted flow at a source node does not only consume the source's bandwidth, but the bandwidth of all the neighboring nodes along the data propagation path, thereby affecting ongoing flows of other nodes. Hence, it is essential to perform admission control along the entire path.

Flow Reservation

The resource reservation arranges for the allocation of suitable end-system and network resources to satisfy the user QoS specification. In doing so, the resource reservation interacts with the QoS routing protocol to establish a path through the network in the first instance, then, based on admission control at each node, end-to-end resources are allocated.

RSVP (Resource Reservation Setup Protocol) [4] is a signaling mechanism to carry the QoS parameters from the sender to the receiver to make resource reservations along the path. The mechanism works as follows:

- The sender of an application sends PATH messages containing the traffic specifications to the receiver(s) of the application that will use this reservation.
- The receiver receives this PATH message and sends RESV message to the sender specifying the flow it wants to receive.
- As the RESV message flows back to the sender, reservations are made at every node along the way. If at any point along the path the request cannot be supported, that request is blocked.

- At every router/host along the way, path and reservation states are maintained for every application session. Periodically sent PATH and RESV messages refresh the path and reservation states.

RSVP is designed to provide integrated service for packet-switched network such as IEEE802.3. However, because of the scarcity of bandwidth and high link error in wireless network, directly applying RSVP may lead to high overhead and instable performance.

Description of AODV

The Ad Hoc On-Demand Distance Vector (AODV) routing protocol [20], enables multi-hop routing between participating mobile nodes wishing to establish an ad-hoc network. It's basically a combination of DSDV and DSR. It borrows the basic on-demand mechanism of Route Discovery and Route Maintenance from DSR, plus the use of hop-by-hop routing, sequence numbers, and periodic beacons from DSDV. AODV minimizes the number of required broadcasts by creating routes on an on-demand basis, as opposed to maintaining a complete list of routes as in the DSDV algorithm.

Features of this protocol include loop freedom and that link breakage cause immediate notifications to be sent to the affected set of nodes, but only that set. It uses destination sequence numbers to guarantee freshness of a route.

The algorithm uses different messages to discover and maintain links. Whenever a node wants to try and find a route to another node, it broadcast a Route Request (RREQ) to all its neighbors. The RREQ propagates through the network until it reaches the destination or a node with a fresh enough route to the destination. Then the route is made available by unicasting a Route Reply (RREP) back to the source.

The algorithm uses HELLO messages (a special RREP) that are broadcasted periodically to the immediate neighbors. These HELLO messages are local advertisement for continued presence of the node and neighbor using routes through the broadcasting node will continue to mark the routes as valid.

Hello Messages

Another aspect of the AODV protocol is the use of HELLO messages, periodic local broadcasts by a node to inform each mobile node of other nodes in its neighborhood. Hello messages can be used to maintain the local connectivity of a node. Nodes listen for retransmissions of data packets to ensure the next hop is still within reach. If such a retransmission is not heard, the node may use any one of a number of techniques, including the reception of HELLO messages, to determine whether the next hop is within communication range. The HELLO messages may list the other nodes from which a mobile has heard, thereby yielding a greater knowledge of the network connectivity

Methodology

In our approach we are utilizing HELLO messages from AODV routing protocol to send bandwidth information to neighbors, so that they can make necessary reservations based on the available bandwidth. Our approach tries to solve the problem of determining interference caused by transmission between two nodes in an 802.11 ad-hoc network in other nodes that are in their coverage area.

We also highlight the problems of differentiation mechanisms. In order to solve these issues we first carry out an evaluation of service differentiation mechanisms by way of simulations. On tackling the problem the problem, there is a need for nodes [1](stations) to be equipped with the following:

- Resource estimation (estimating available bandwidth).
- Admission control based on available bandwidth.

[1] Note that we use node and station interchangeably in this paper.

- Flow reservation after the admission control.

These three factors are fundamental in greatly reducing the problems of traffic degradation as load increases in the network.

Resource Estimation

In a distributed ad hoc network, a host's available bandwidth is not only decided by the raw channel bandwidth, but also by its neighbor's bandwidth usage and interference caused by other sources, each of which reduces a host's available bandwidth for transmitting data. Therefore, applications cannot properly optimize their coding rate without knowledge of the status of the entire network. Bandwidth is a fundamental resource. When flows are routed through the network, estimating the remaining bandwidth is often required before performing admission control, flow management, congestion control or routing based on bandwidth constraints.

Bandwidth estimation can be done using various methods; for example, in [27] bandwidth estimation is a cross-layer design of the routing and MAC layers, and in [14], the available bandwidth is estimated in the MAC layer and is sent to the routing layer for admission control. Therefore, bandwidth estimation can be performed in several different network layers. We are using a similar approach described in [27].

To determine whether there is enough bandwidth available for a new flow, all we need to know is the available link capacity and the bandwidth to be consumed by the requesting flow. In wired networks this is a trivial task since the underlying medium is a dedicated point-to-point link with fixed capability.

However, in wireless networks the radio channel of each node is shared with all its neighbors. Because of the shared medium, a node can successfully use the channel only when all its neighbors do not transmit and receive packets at the same time. We call this the aggregation effect.

In [5] the authors derive formulae to estimate the available bandwidth in an ad-hoc network using shared links. To do so, each node may do the following calculation:

$$MUBi = Ci - \sum lij, \forall j \in Neighbourhood\ \ of\ \ i$$

$MUBi$ is the maximum unused bandwidth, Ci is the capacity of the node and lij is the total traffic between nodes *i* and *j*.

But, since the traffic between neighbors of a node also interfere; these traffics must also be taken into consideration to calculate the maximum available bandwidth (*MABi*), what leads us to:

$$MABi = MUBi - \sum_j \sum_k ljk, \forall j \in Neighborhood - of - i, \forall k \in Neighborhood - of - j$$

Using Hello messages from AODV routing protocols, all nodes can broadcast their *MUB* and their local bandwidth requests. This makes sure that all nodes are aware of their neighbors' traffic demands.

When using a reactive routing protocol, such as AODV, the *MAB* may be used to elect a path that fulfills the QoS needs of a flow. The Route Request (RREQ) messages checks the available bandwidth to be sure that the flow may pass through the node (if not, the RREQ is discarded). During the reverse path establishment (Route Reply), the resources may be then reserved.

The previous formulas, however, may not guarantee a correct calculation of the available bandwidth in the general case. Available bandwidth can be computed if the nodes know not only lij, but the *MABi* computed by their neighbors. Then, the available bandwidth ABi to allocate new reservations at $Node_i$ is given by:

$$ABi = \min\{MABi, MABj\}, \forall j \in Neighborhood - of - i$$

Admission Control

Service differentiation is helpful in providing better QoS for multimedia data traffic under low to medium traffic load conditions. However, due to the inefficiency of IEEE 802.11 MAC, service differentiation does not perform well under high traffic load conditions.

In service differentiation mechanisms, no assurance can be given to higher priority traffic in terms of throughput and delay performance. Admission control is an important tool to maintain QoS experienced by users. By predicting the achievable throughput of data flows and avoiding channel overloading, the QoS of existing flows can be maintained. Admission control is based on local computation of the available bandwidth by each node of the network based on information that is sent by its neighbors through periodical HELLO messages.

Flow Reservation

The resource reservation arranges for the allocation of suitable end-system and network resources to satisfy the user QoS specification. In doing so, the resource reservation interacts with the QoS routing protocol to establish a path through the network in the first instance, then, based on admission control at each node, end-to-end resources are allocated.

For in-band signaling protocols for MANET such as INSIGNIA [15], the reservation control message is integrated together with the data packet. In our approach we used HELLO messages which are extended to include bandwidth field, which carries bandwidth information from neighbors. Reservation Request and Reply messages are integrated in AODV as described in [21]: the bandwidth reservation is included in a Route Request (RREQ) message as an extension object. The RREQ QoS extensions include a session-ID to identify the flows together with the Source and Destination addresses.

Upon receiving a RREQ, intermediate nodes apply the admission control algorithm. If the reservation is accepted, the RREQ is forwarded, and it is discarded otherwise. However, reservation is only done when the RREP is received (see Figure 5). Opposite to AODV, if an intermediate node has a route to a destination, this node should not answer with a route reply to the sender, since the intermediate node does not know whether further nodes can accomplish the bandwidth reservation. In order to avoid this situation the D flag of a RREQ is activated [19], indicating that only the destination can send a RREP.

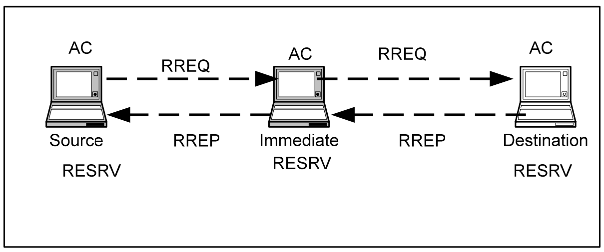

Figure 5. Reservation Procedure

Tools Used

Network Simulator (NS-2)

Network Simulator 2 (NS-2) [18] is a simulation tool originated from Lawrence Berkeley National Laboratory. We used NS-2 to provide substantial support for our simulation of routing and multicast protocols over wired and wireless networks. NS-2 has an advanced 802.11 module, which is applied and verified extensively in the network community, hence its an excellent simulation tool within this research.

Xgraph and Gnuplot

Xgraph [25] and Gnuplot [12] are X-Window applications that include interactive plotting and graphing, and animation and derivatives. We used the programs to create graphic representations of simulation results. Output data from TCL scripts are used as data sets to Xgraph or Gnuplot. (See Figure 6)

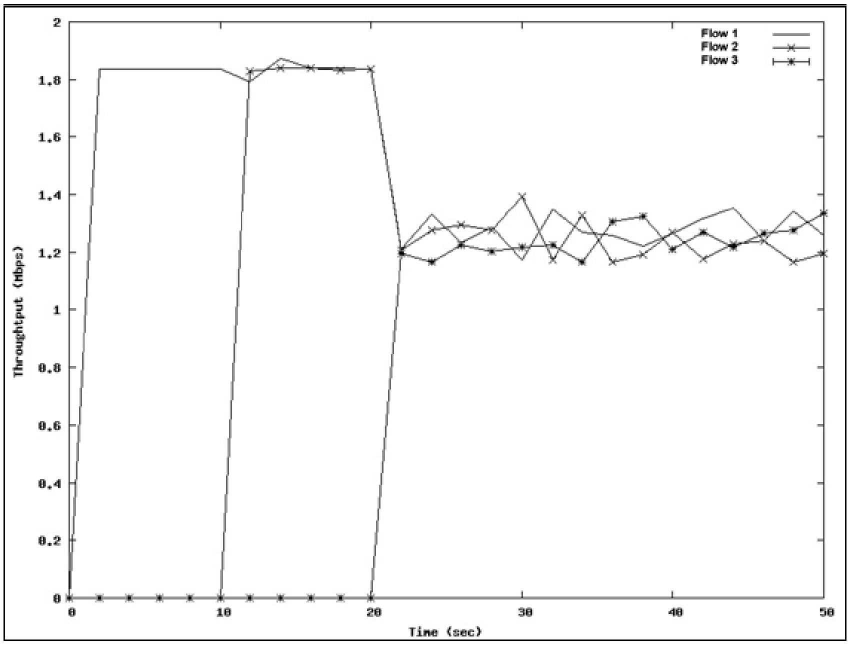

Figure 6. Gnuplot running comparing three trace files in a graph

Other tools

Other tools include AWK, *grep* and Perl scripts; these are mainly used to extract important statistics information from trace files. AWK utility allows us to do simple operations on data files such as averaging the values of a given column, summing or multiplying term by term between several columns. In our work we extensively used this utility to calculate and extract QoS metrics from trace files. The *grep* command in UNIX allows to "filter" a file. This is important because some generated trace files are enormous hence needs to be filtered.

Simulation Results

The main aim is to analyze the differentiation levels offered by modifying MAC parameters, such as DIFS and contention window, in different scenarios and then expose shortcomings from these mechanisms. We use these parameters because they are very similar to 802.11e differentiation mechanism. The Network Simulator NS-2 [18] version 2.27 is used in the simulations studies. We have used the functionalities of 802.11 networks added with service differentiation, ad hoc routing (AODV) and Constant Bit Rate (CBR) traffic source.

Simulation Environment

In this Section we describe the environment and the scenarios used in our simulations.

First Scenario

The topology we used in our first experiment consists of three stations transmitting to a fourth station. All stations generate 1.8 Mbps CBR traffic with packet sizes of 1000 bytes. The distance between stations is 50 meters. The ad-hoc routing protocol is AODV and the channel capacity is 11Mbps. The maximum achievable throughput in this channel is largely dependent on the frame size used by the sources, the use of the RTS/CTS handshake, and the number of stations contending for the medium, and the differentiation parameters such as DIFS and the contention window. For example, when only one CBR source contends for the medium and uses packet of 1000 bytes, the maximum achievable throughput is about 3.6 Mbps. Stations 1,2 and 3 start their transmissions at 0s, 10s and 20s to station 0. The throughput obtained by each station at each 1 s interval is evaluated. (See Figure 7)

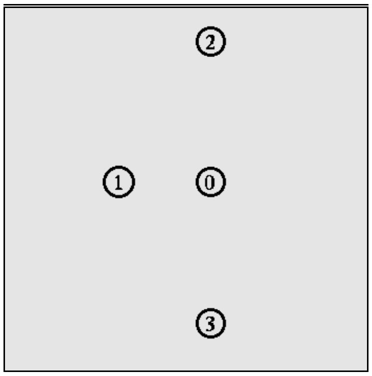

Figure7. First Simulation Topology

First Simulation: No Differentiation Mechanism

In the first simulation, no differentiation mechanism is used, i.e., all nodes have the default configuration for differentiation parameters, as show in the Table 1.

Table 1. Default Configuration Parameters for 802.11 DCF

Parameter	**Value**
DIFS	50us
CWmin	31
CWmax	1023

DIFS-based Scheme

Our second simulation is that of DIFS-based scheme in which stations are given different DIFS parameters. Here we had to make changes to NS-2 to enable it expose the DIFS parameter so that each node uses its own value and then recompile it. Parameters used are in Table 2.

Table 2. DIFS parameter for DIFS-based Scheme

Stations	**DIFS**	**CWmin :CWmax**
1	20us	31:1023
2	40us	31:1023
3	60us	31:1023

Backoff-based scheme

Our third simulation is that of changing the contention window size and then study the behavior of throughput. Each station is configured with a contention window interval [CWmin: CWmax] as show in Table 3. Stations with smaller intervals have higher priority of accessing the channel compared to stations with larger values. So we expect stations with smaller interval values to have higher throughput compared to those with larger values.

Table 3. Contention Window Sizes

Station	CW (CWmin : CWmax)	DIFS
1	31 -1023	20us
2	63 -2047	20us
3	127 - 4095	20us

<u>Second Scenario: Increasing the number of stations</u>

In this scenario we study the level of differentiation and how throughput behaves as the number of stations contending for the wireless channel increases. We are still using three types of priorities (classes) for both DIFS-based scheme and backoff-based scheme. All stations send traffic to one station and we maintain the same configuration parameters as stated in First Simulation with no differentiation mechanism. The number of stations per priority class is increased from 3 to 10.

<u>Third Scenario: Introducing Mobility</u>

In our third scenario, we introduced some mobility so that we may understand the influence of mobility on differentiation and throughput. In this scenario, all stations follow a random generated movement pattern, called random-way point [13], in which transmitting stations always stay within the range of the receiving station. Average speed of each station is 10 m/s with movement pauses of 20 seconds in average. The number of stations is increased in each simulation run as described in second scenario. The movement pattern is generated by a utility found in NS-2 called *setdest.*

Results

In this Section we provide the results obtained from the different simulation scenarios described above. All graphs were obtained using Gnuplot and setting its terminal to jpeg.

No Differentiation (Pure DCF)

This first simulation shows the throughput of three CBR flows over DCF without any differentiation. This clearly shows how DCF is best-effort service in which each station is given equal bandwidth. Figure 8 shows the first station starts to transmit at 0 seconds and gets its required bandwidth (1.8Mbps), and then station 2 joins in after 10 seconds, it also gets its required bandwidth. But problems start after 20 seconds when station 3 transmits, the bandwidth reduces to about 1.2 Mbps for each station. The introduction of the third flow causes the total traffic to go up to 5.4Mbps (3 x 1.8 = 5.4Mbps), this is more than the maximum available bandwidth of about 3.6 Mbps (this is calculated according to the formula suggested in [8]).

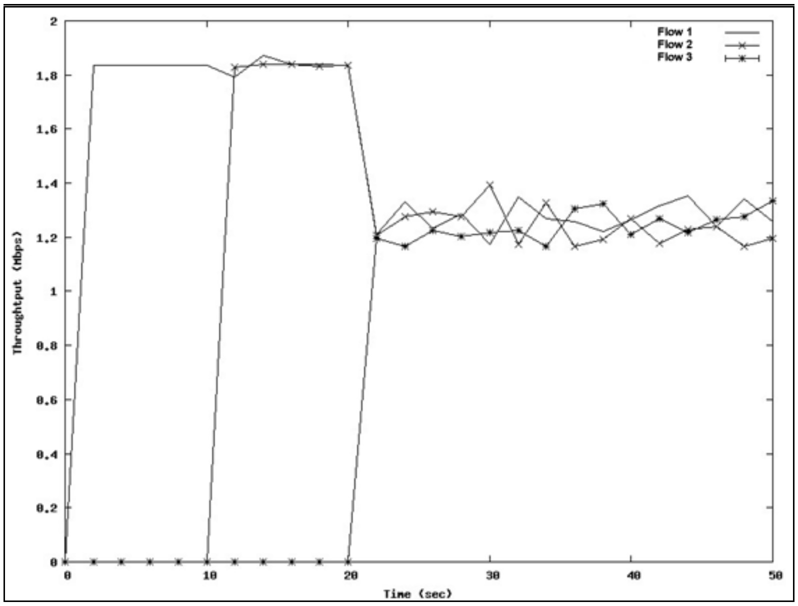

Figure 8. Throughput of 3 flows with No Differentiation

<u>DIFS –based Differentiation</u>

In this Section we provide our results for DIFS-based differentiation.

First Scenario

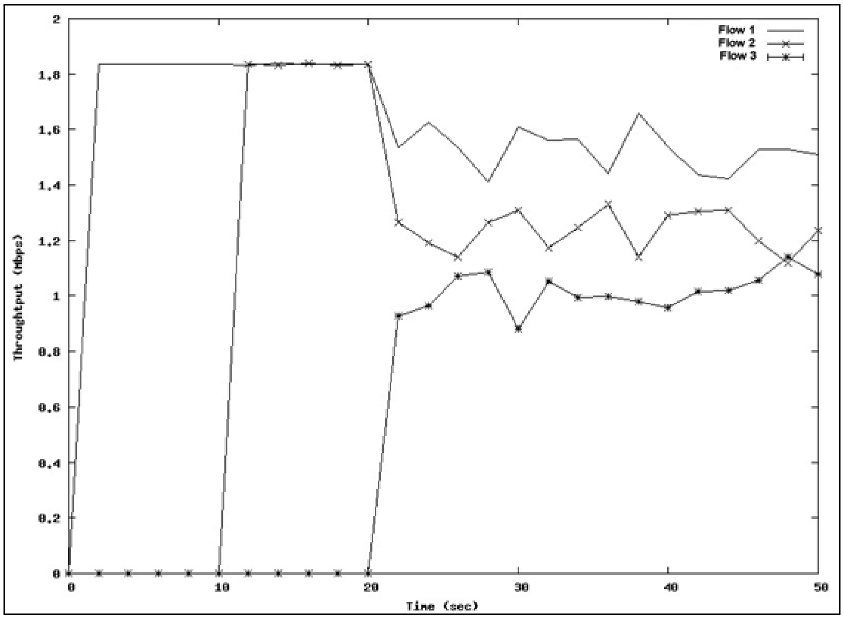

Figure 9. Throughput of 3 flows with DIFS-based Differentiation

Figure 9 shows the throughput differentiation of DIFS-based scheme. Between 10 and 20 seconds, only two stations fairly share the channel because their aggregate rate is inferior to the maximum achievable throughput. When the third station starts transmitting, the channel capacity is lower than the total traffic and differentiation starts. Station 1 (Flow 1) obtains more bandwidth than stations 2 (Flow 2) and 3 (Flow 3), because it has the smallest DIFS.

Second Scenario: Increasing Number of Nodes

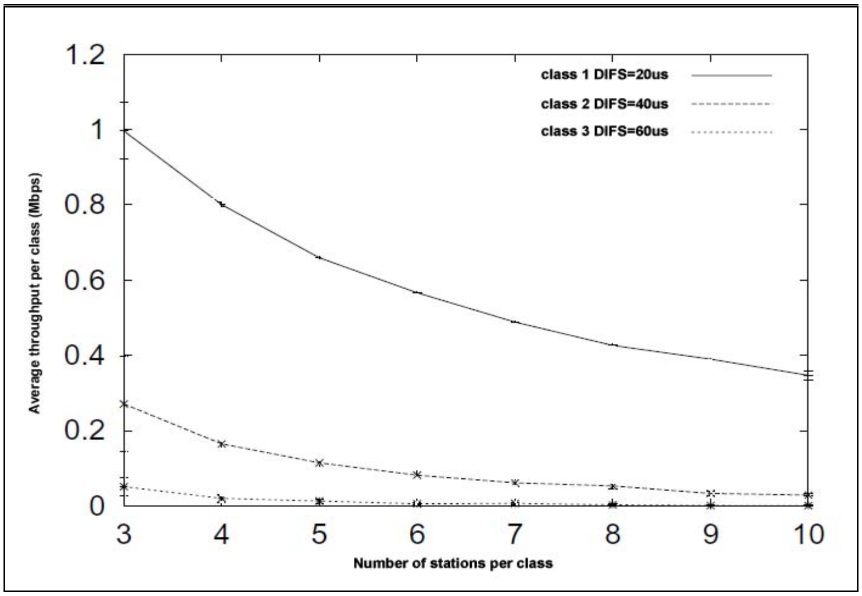

Figure 10. DIFS: Average throughput when increasing the number stations per class

Figure 10 shows the average throughput of stations that belong to one of the service classes, the same throughput differentiation is achieved with the presence of an increasing number of stations. However the throughput decreases with increase in the number of stations. This is mainly attributed to the shared nature of wireless channels.

Third Scenario: Introducing Mobility

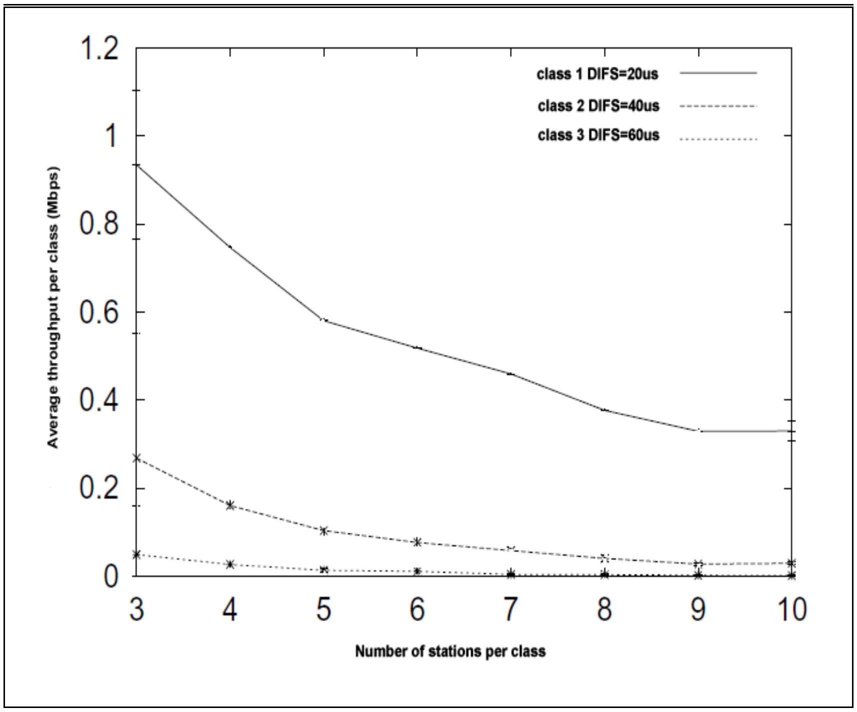

Figure 11. DIFS: Average throughput under Mobility

Figure 11 shows the average throughput obtained by stations within the same service class. Results show that the throughput differentiation takes effect even in the presence of mobility.

Backoff-Based Differentiation

In this Section we provide our results obtained for backoff-based differentiation.

First Scenario

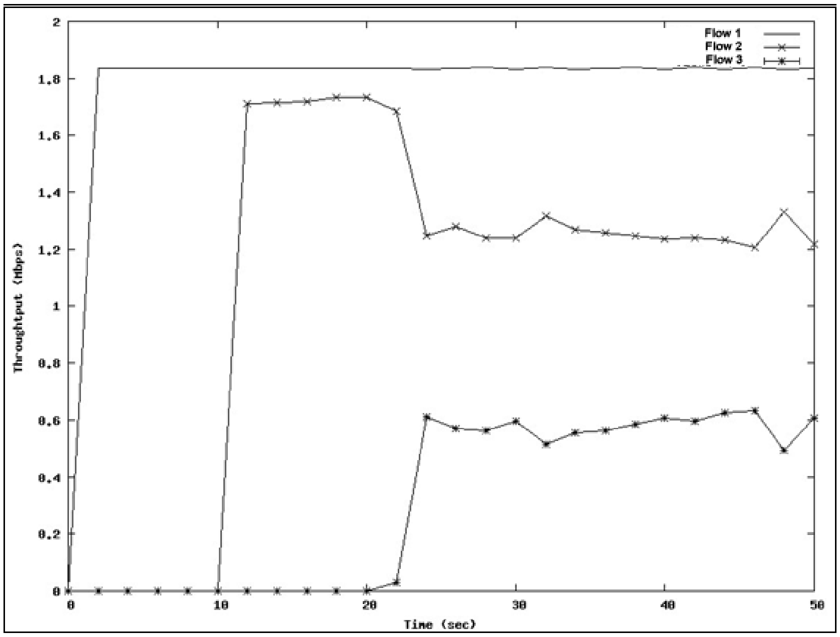

Figure 12. Throughput for Backoff -based Differentiation

Figure 12 shows clearly that the station (flow 1) with a smaller contention window interval have more bandwidth. Station 2 (flow 2) and 3 (flow 3) have their bandwidth decreased, hence making them share. However, station 2 still has more bandwidth compared to 3 because it has a smaller contention window. The results clearly show that stations with a smaller CWmin value obtain a larger share of the channel capacity than the other stations.

Second Scenario: Increasing the Number of Nodes

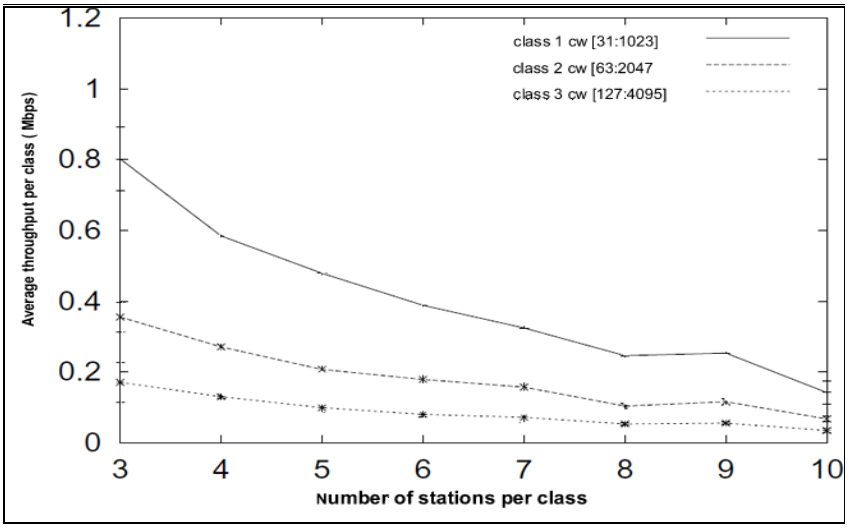

Figure 13. Backoff-Based: Average throughput when increasing the number of stations per class

Figure 13 shows the average throughput of stations that belong to one of the service classes, the same throughput differentiation is achieved with the presence of an increasing number of stations. However the throughput decreases with increase in the number of stations.

Third Scenario: Introducing Mobility

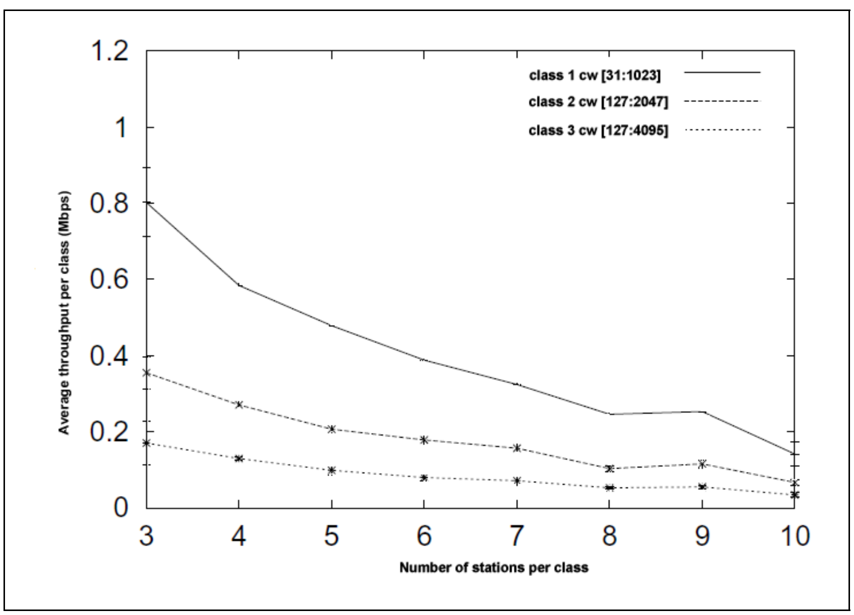

Figure 14. Backoff-based: Average throughput under mobility.

Figure 14 shows the average throughput obtained by stations within the same service class. Results show that the throughput differentiation takes effect even in the presence of mobility.

Observations and Analysis

In this Section we provide our observations and analysis of obtained results.

Backoff-based Scheme

If the total number of stations increases, the throughput per node decreases rapidly. The total amount of effective channel capacity drop due to increased number of collisions, and the decreased channel capacity is shared among a larger number of stations.

Stations with a smaller CW value have a higher probability to transmit. Therefore, if the other parameters are set equal, stations with a smaller CW value will have a larger share of the medium capacity. This explains the reasons why stations with smaller contention window interval have higher throughput.

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

CWmax also contributes to the evolution of the CW in the contending process. Stations with smaller CWmax values are expected to obtain a larger share of the medium capacity. However, as opposed to CWmin, the value of CWmax is only reached after a number of successive collisions with the same packet. This suggests that differentiated CWmax values will only have effect with a relatively high number of contending stations. When a number is “relatively high” depends on the exact values of CWmin and CWmax, together with other parameters. CWmax defines the final CW value of its exponential growing process. Therefore, the value of CWmax is even more critical if collisions take place frequently, such as in a network with a high number of contending stations. A small CWmax value can greatly downgrade the system performance on channel efficiency if the number of contending stations is high.

DIFS-based scheme

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

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

This scheme may suffer from inefficient channel usage since even if the majority of the traffic is from the class with the larger DIFS, they all must wait a very long time before they can compete for the channel [2].

The impact of MAC parameters on the throughput is shown in Table 4.

Table 4. Impact of MAC parameters on throughput

Parameters	**Effect on Channel Efficiency (throughput)**	
	Small number of contending stations	Large number of contending stations
CWmin	Slightly positive	Significantly negative
CWmax	Hardly present	Significantly negative
DIFS	Hardly present	Positive

Evaluating the Impact of HELLO messages

In order to allow each node to have the complete set of information that is needed in order to compute its available bandwidth, some values must be piggybacked in HELLO messages. Since the amount of information to be transmitted by each node is proportional to the number of neighbors of the node, it is important to evaluate the overhead caused by this extra information.

μi	***xij***	**addrj**	**---------**

For very neighbor *j* of *i*

Figure 15. Information needed to be carried by HELLO messages

Figure 15 shows the extra information required to be sent on each HELLO message in order to enable nodes compute *B* by using the following:

$$Bi = Q - \left(\mu i + \sum_{j \in Ni} \mu j + \sum_{j \in Ni, k \notin Ni+} \chi jk \right)$$

Considering the case that we represent allocated bandwidth (both μi and *xjk*) as variables with β bits of length, neighbors address as variables with α bits of length and that a node has N neighbors, the overall overhead in HELLO messages (Ω) may be computed as:

$$\Omega = \beta + N \times (\alpha + \beta) \quad \textbf{(In bits)}$$

If we consider that HELLO messages are broadcasted and that RTS/CTS are not used on broadcast messages, each node will see its total traffic increased, not only due to this extra information sent by its own HELLO messages; but also by the ones sent by its neighbors. Hence, the total overhead ($\Omega total$), caused by this extra information in each node, can be computed as:

$$\Omega total = \Omega \times (N + 1)$$

$$\Omega total = [\beta + N \times (\alpha + \beta)] \times (N + 1)$$
$$\Omega total = (\beta + \alpha) N^2 + (\alpha + 2\beta) N + \beta$$

If we use the full IP address as the node identification, the overhead will be extremely high (32 bits for IPv4 and 128 bits for IPv6 per neighbor), instead, we may use a hash of its IP address and represent each node with only 8 bits (the use of an appropriate hash function may provide a very low probability of mistakes).For the same reason, it is not very efficient to represent reserved bandwidths (both μi and χjk) in bps. As an alternative, these values may be represented as multiples of 8Kbps (which would be the minimum reserveable bandwidth). By using this approach and limiting the maximum reservable bandwidth of a flow to 2Mbps, we may represent reserved bandwidths with only 8 bits as well. Other combinations of minimum and maximum reservable bandwidths may also be obtained by using 8 bits to represent a reservation, although the one used as an example seems to be reasonable.

By using the proposed lengths for representing values to be transmitted, the following additional bits will be transmitted by each HELLO message:

$$\Omega total = (16)N^2 + (24)N + 8$$

Figure 16 shows the impact of extending HELLO messages in the traffic seen by a single node. Although the growth is linear, HELLO packets of the node also see every neighbor, causing the overall traffic to grow exponentially. However, although the traffic growth is exponential, considering that AODV RFC [19], recommends a HELLO interval of 1 second, even when the networks is dense (20 neighbors), the extensions in the HELLO messages causes traffic in one node to increase below 7Kbps.

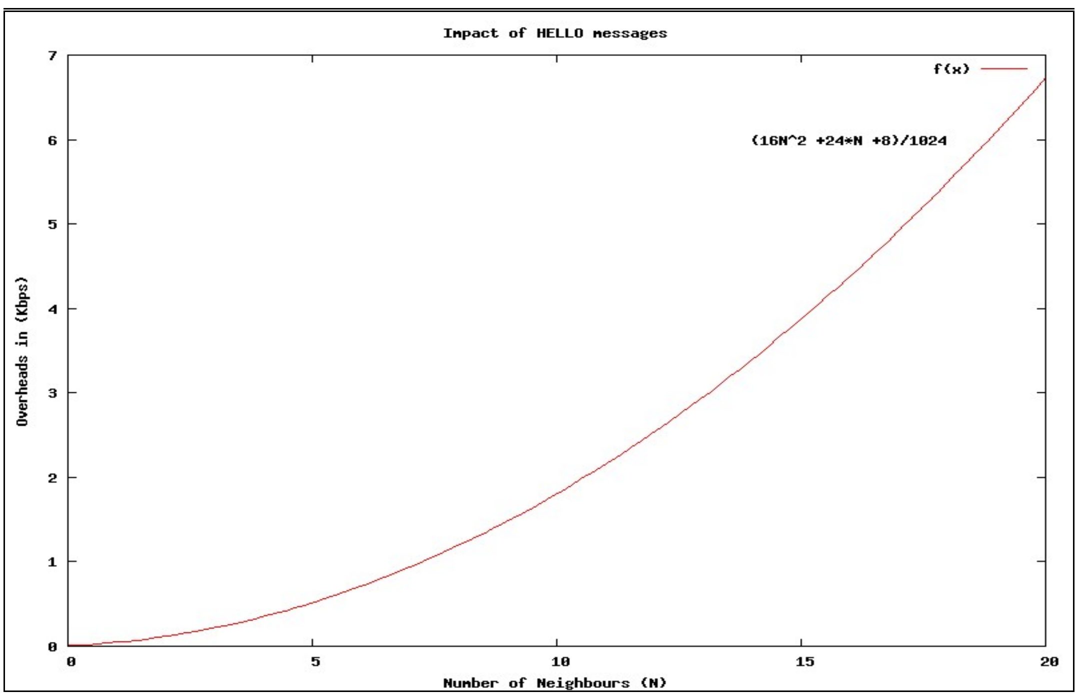

Figure 16. Impact of extra information being carried by HELLO messages

CONCLUSION

From our results we concluded that differentiation service mechanisms at the 802.11 MAC layer managed to provide quality of service under low network load; but as the traffic increases, performance of these mechanisms perform poorly. They do not also provide any admission control mechanism in order to control the amount of traffic injected into the network. Hence, without admission control and resource allocation, providing QoS guarantees by only differentiating flows and coordinating the order of channel access, cannot be effective under high traffic loads. After evaluating the impact of HELLO messages, we noticed that the traffic impact of adding these extra data in HELLO packets is very low; even for dense ad-hoc networks.

Zimbabwe Telecommunications industry can use MANETS in boosting communication and move with technology. MANET are cheap to run, there is no central administration, less physical infrastructure as opposed fixed wired networks.

For the future, we plan to present QoS reservation strategy that will take into account the issues of interference from neighborhood traffic in order to compute its available bandwidth.

References

Aad, I. and Casterlluccia, C. Remarks on per-flow differentiation in IEEE 802.11. In Proceedings of the European Wireless, Florence, Italy, February 2002.

Aad, I. and Casterlluccia, C. Differentiation mechanisms for IEEE 802.11. Infocom '2001, Anchorage, Alaska, April 2001.

Ahn, G.S., Campbell, A., Veres, A., and Sun, L.H., Supporting Service Differentiation for Real-Time and Best-Effort Traffic in Stateless Wireless Ad Hoc Networks (SWAN). IEEE Transactions on Mobile Computing, 1(3):192–207, July-September 2002.

Braden, R., et al. RFC 2205: Resource ReSerVation Protocol (RSVP), Sep 1997.

Cansever, Derya H., Arnold M. Michelson, and Allen H. Levesque. Quality of service support in mobile ad-hoc IP networks. In MILCOM 1999 - IEEE Military Communications Conference, number 1, pages 30.34, Oct 1999.

Corson, S., and Macker, "Mobile Ad-hoc Networks: Routing Protocol Performance Issues and Evaluation Considerations", IETF RFC-2501, January 1999

IEEE 802.11e WG. .Draft supplement to Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications: MAC Enhancements for Quality of Service (QoS), IEEE Std 802.11e/D9.0.August 2004.

Jun Jangeun, Pushkin Peddabachagari and Mihail Sichitiu, "Theoretical Maximum Throughput of IEEE 802.11 and its Applications," in Proc. of the 2nd IEEE International Symposium on Network Computing and Applications (NCA'03), Cambridge, MA, April 16-18, 2003, pp. 249-256

Kai Chen, Samarth, H. S., and Klara Nahrstedt, "Dynamic Bandwidth Management for Single-hop Ad Hoc Wireless Networks". In proceedings of 1st IEEE International Conference on Pervasive Computing and Communications, Dallas-Fort Worth, Texas, pp. 195-203, March 2003.

Kanodia, V., Li, C., Sabharwal, A., Sadeghi, B., and Knightly, E., Distributed multi-hop scheduling and medium access with delay and throughput constraints. In ACM The Seventh Annual Conference on Mobile Computing and Networking (MobiCom), Rome, Italy, July 2001

Kanodia, V., Chengzhi Li, A. Sabharwal, B. Sadeghi, and E. Knightly. Distributed priority scheduling and medium access in ad-hoc networks. Wireless Networks,8(5):455-466, September 2002

Kelley Colin, Thomas Williams, and others, "Gnuplot Home page", <URL: http://www.gnuplot.info/faq/>, Accessed Nov 2005.

Kevin Fall and Kannan Varadhan. The NS Manual. Technical report, The VINT Project, January 2002

Kunz, T., Ge Y., and Lamont, L., "Quality of Service Routing in Ad-hoc Networks Using OLSR," in Proceeding of the 36th Hawaii International Conference on System Science, 2003.

Lee, S.B. and Campbell, A.T. “INSIGNIA: In-band signaling support for QoS in mobile ad-hoc networks”. In 5th Int. Workshop on Mobile Multimedia Comm. (MoMuc’98), Berlin, Germany, October 1998.

Leung, R., Liu, J., Poon, E., Chan, A. and Li, B., “A QoS-aware Multi-path Dynamic Source Routing Protocol for Ad-hoc Networks”, in the Proceedings of the 26th IEEE Annual Conference on Local Computer Networks (LCN-1), pp. 132-141, Tampa FL-USA November 2001.

Nair, R., Crawley, E., Rajagopalan, B., and Sandick, H., “A framework for QoS based routing in the internet,” Aug. 1998, RFC 2386.

Network Simulator 2 (NS-2) Homepage. <URL:http://www.isi.edu/nsnam/ns/ > Accessed September 2005.

Perkins, C. E. and Belding-Royer, E., Requests For Comments (RFC3461). <URL:http://www.ietf.org/internet-drafts/draft-ietf-manet-aodv-08.txt> Accessed November 2005.

Perkins, C. E. and Royer, E. M., Ad-hoc On-Demand Distance Vector Routing. In Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications, pages 90–100, New Orleans, LA, February 1999.

Perkins, C., and Belding-Royer E. Quality of Service for Ad hoc On-Demand Distance Vector Routing (work in progress), Oct 2003. draft-perkins-manet-aodvqos-02.txt

Qiang Ni, Lamia Romdhani, and Thierry Turletti, “Adaptive EDCF: Enhanced Service Differentiation for IEEE 802.11 Wireless Ad Hoc Networks”. IEEE Wireless Communications and Networking, New Orleans, USA,Vol. 2, pp. 1373 –1378, March 2003.

Society, I.C. IEEE standard for Wireless LAN-Medium Access Control and Physical Layer Specification, November 1999.

Wang Yu and Bensaou Brahim. Achieving fairness in IEEE 802.11 DFWMAC with variable packets lengths. In IEEE Global Telecommunications Conference (Globecom), Texas, USA, November 2001.

XGraph homepage <URL:http://www.isi.edu/nsnam/xgraph/ >, Accessed November 2005

Xiao, H., Seah, W.K.G., A. Lo and K.C. Chua, 2000.A flexible quality of service model for mobile ad-hoc networks, IEEE VTC2000-spring, Tokyo, Japan, May, 2000.

Xue, Q. and Ganz, A., “Ad hoc QoS On-demand Routing (AQOR) in Mobile Ad hoc Networks,” in Journal of Parallel and Distributed Computing (JPDC), February 2003.

Zhu, H., Li, M., Chlamta, I. and Prabhakaran, B., “A Survey of Quality of Service in IEEE 802.11 Networks,” IEEE Wireless Communications, August 2004.