Towards End-to-End QoS in Ad Hoc Networks

Osman Salem, Abdelmalek Benzekri

Towards End-to-End QoS in Ad Hoc Networks

OSMAN SALEM and ABDELMALEK BENZEKRI
Institut de Recherche en Informatique de Toulouse,
Université Paul Sabatier,
118 Route de Narbonne - 31062 Toulouse Cedex 04 - France
Email:{benzekri, osman}@irit.fr

***Abstract*— In this paper, we address the problem of supporting adaptive QoS resource management in mobile ad hoc networks, by proposing an efficient model for providing proportional end-to-end QoS between classes. The effectiveness of our proposed solution in meeting desired QoS differentiation at a specific node and from end-to-end are assessed by simulation using a queueing network model implemented in QNAP. The experiments results show that the proposed solution provides consistent proportional differentiation for any service class and validates our claim even under bursty traffic and fading channel conditions.**

I. Introduction

A mobile Ad Hoc network (MANET[1]) is a collection of autonomous mobile hosts, where each one is equipped with a wireless card that makes it able to communicate with any other host, directly if this last is in the same receiving zone, or indirectly through intermediate hosts that forward packets towards the required destination. Therefore, each host acts as a router when cooperating to forward packets for others, as well as a communication end point.

With the evolution of wireless communications and the emergence of diversified multimedia technologies, quality of service in ad hoc networks became an area of great interest. Besides existing problems for QoS in IP networks, the characteristics of MANETs impose new constraints due to the dynamic behavior of each host and the variation of limited available resources.

A lot of researches has been investigated in routing area [1], [2], [3], [4], [5], [6], [7], and today routing protocols are considered mature enough to face energy constraints and frequently changing network topology caused by mobility (e.g., DSR [2], [6], AODV [3], [6], etc.). Many QoS aware routing protocols that claim to provide a partial (or complete) solution to QoS routing problems have appeared consequently, e.g. QoS-AODV [4], MP-DSR [5], ASAP [8], CEDAR [9].

In the current days, the Integrated services (IntServ) [10] and the differentiated services (DiffServ) [11] are the two principal architectures proposed to provide QoS in wired networks. While the IntServ approach achieves end-to-end services guarantees through per-flow resource reservations, the DiffServ focuses on traffic aggregates and provides more scalable architecture. Contrarily to IntServ, DiffServ does not require any per-flow admission control or signaling, and routers do not maintain a per-flow state information. Routers in DiffServ domain, need only to implement a priority scheduling and buffering mechanism in order to serve packets according to specified fields in their headers.

The migration of these architectures to MANETs is proved to be inconsistent with the characteristics of these networks [12], [13]. Many researches have been based on these concepts and the mitigation of their impediments to make them suitable with the characteristics of MANETs, like INSIGNA [14], [15], FQMM [16], and SWAN [17]. Most of these strategies rely on admission control, priority based resource allocation and scheduling. They only ensure that a newly added flow achieves its desired QoS but can not prevent the degradation of existing flows due to the contention with newly admitted flows and link breaking.

Taking into account that the bandwidth fluctuations and routes change over time, existing QoS mechanisms that require explicit resource reservation and absolute QoS guarantee are difficult to realize in ad hoc networks. The mobility and link breaking make useless the network resources reservation to provide hard guarantee when the resources do not exist. Clearly, there are a number of reasons to beleive that per-hop differentiation technique is more appropriate for ad hoc networks than resources reservation. Without loss of generality, DiffServ is addressed to the network core by aggregating flows in a set of classes. Thus allows per-hop differentiated services for the aggregated flows in the core of the network without requiring any resources allocation. However, DiffServ architecture largely depend on available resources and it does not define any scheme for taking corrective actions when congestion occurs. This is why a static DiffServ model is not suitable for ad hoc networks, and it is imperative to use some kind of feedback as a measure of the conditions of the network to dynamically regulate the class of traffic in the network with respect to the perceived and required QoS.

We turn our attention to provide an adaptive approach with a soft guarantee (small time scale violation) rather than absolute one (strict guarantee). Adaptive services are very attractive in ad hoc networks, because networks resources are relatively scarces and widely variables, where resources fluctuations are mostly caused by mobility, energy constraints and channel fading. In contrast to traditional techniques (IntServ and DiffServ), our model adjusts the spacing of QoS perceived by each class proportionally and independently of network load.

Our approach to provide proportional end-to-end QoS in ad hoc is based on the idea of fighting QoS degradation due to mobility, which will change the topology and will produce bandwidth fluctuation due to load redistribution when re-routing existing traffic, after link break caused by node movement outside the radio range of its vicinity. Therefore, the relayed packets by this node of an active communications flow towards a destination will be inevitably lost. To recover communication, source node initiates end-to-end alternative route discovery with reactive routing protocol, and flows will travel through the newly discovered route if there is one, causing load redistribution and changing the QoS perceived by existing traffic profile in the newly traversed route due to additional amount of traffic.

The Proportional Differentiated Service (PDS) [18], [19] model, which was proposed for IP networks, classifies flows into N classes where class i gets better proportional performance than class $i-1$. PDS aims to achieve better performance for high priority class relatively to low priority class within fixed pre-specified quality spacing. This proportionality is achieved through the use of a scheduling mechanism able to provide the pre-specified spacing between classes. The most important idea of PDS is that even the actual quality of each class will vary with network load, the spacing ratio between classes will remain constant.

A proportional approach outperforms the strict prioritization scheme, where higher priority classes are serviced before lower ones. This means if the high priority classes are persistently backlogged in corresponding queues, the low priority will starve for bandwidth with strict priority scheduling. Moreover, strict priority schemes do not provide a tuning mechanism for adjusting the quality spacing among classes, and the QoS perceived by a class depends only on the load distribution. Another advantage for using proportional differentiated service rather than strict one is that bandwidth degradation is sensed firstly by the low priority flows, whenever it is desirable to distribute the bandwidth degradation across different classes in a proportional fair manner. Furthermore, while some research in ad hoc try to provide a static fairness in resource network distribution, obsessive fairness is neither reasonable nor desirable in this kind of networks, where some applications expect better services than others.

Usually, QoS parameters are specified in term of maximum end-to-end delay, maximum loss rate and minimum throughput. The differentiation between classes in a static manner will not to be able to respond to these end-to-end requirements with resources fluctuations. To resolve this problem, we will use a priority adaptor mechanism (or dynamic class selection mechanism proposed in [20]) to dynamically adjust the priority of each flow according to the perceived/required QoS.

In this paper, we address the problem of providing a proportional differentiation service that supports a wide range variation of the network load in ad hoc networks. The problem of using PDS model lies in the additional random waiting time for every frame due to medium access mechanism (CSMA/CA in IEEE 802.11). Thus render network scheduler inefficient in providing proportionality between classes. Furthermore, the medium access mechanism was initialy proposed to provide bandwidth fairness between contended nodes, but it is unfair to penalize nodes forwarding more traffic for others with a higher delay than its vicinity. To overcome these problems, and to provide consistent delay at all nodes in the path regardless of their arrival rates and their backlogged traffic, our proposed solution uses a dynamic priority adaptor, Waiting Time Priority (WTP) scheduler at the network layer and a contention window adaptor for IEEE 802.11e. The qualitative and quantitative study of our scheme is conducted after a formal description, expressed through stochastic extensions of process algebras and by simulation through a queueing network model. The algebraic description is not described in this paper due to space limitation.

The rest of this paper is organized as follows. Section II gives a brief introduction to the PDS model with the impediments that prevent its use in ad hoc networks. Section III presents the components of our extended PDS (EPDS). Section IV is devoted to the performance evaluation and analysis. Finally, section V concludes the paper with a summary of the results and future directions.

II. THE PROPORTIONAL DIFFERENTIATION SERVICE MODEL AND PROPERTIES

There are two basic types of service differentiation schemes [11]. The first one is the absolute service differentiation, which has a weak ability of adaptation to fluctuating arrival rates from various hosts and which leads to a low resource utilization. The second one is relative service differentiation, where QoS measures for a class are guaranteed relatively to others in the network.

Within the relative service differentiation infrastructure, traffic is divided into N classes that are sorted in an increasing order according to their desired levels of QoS. In this scheme, service quality of class i is better than class $i-1$ for $1 \leq i \leq N$. Thus assures that the class with higher priority will receive a relatively better quality than the classes with lower ones. Therefore, the application must regulate its priority levels to meet its end-to-end requirement in an adaptive manner.

The primary objective of relative service differentiation is to provide proportional differentiated level of QoS to different traffic classes even in the presence of burst in a short timescales. The best known effort in the PDS model was initially proposed in [18], [20], [21], which attempts to provide proportional queueing delay differentiation for packet forwarding in IP networks. It states that the average delay examined by classes should be proportional to the differentiation parameters:

$$\frac{d_i(t,t+\tau)}{d_j(t,t+\tau)} = \frac{\delta_i}{\delta_j},\ \forall\, i \neq j\ and\, i,j \in \{1,2,\ldots,N\} \qquad (1)$$

The class parameters δ_i, δ_j are the pre-specified differentiation parameters for class i and j respectively. They are ordered as that higher classes provide lower delay, i.e. $\delta_1 > \delta_2 > \ldots > \delta_N > 0$. $d_i(t,t+\tau)$, $d_j(t,t+\tau)$ are the average delay for class i, j in the time interval $[t,\ t+\tau]$.The delay perceived by each class is relative to another class, and the higher class will get a better service (i.e. lower delay) than lower classes. Equation 1 must hold for each class regardless of its loads, which mean that the ratio between classes will remain constant depending only on the pre-defined differentiation parameters.

As far as the design for scheduling algorithms to provide proportional delay differentiation, many schedulers have appeared to achieve this proportionality, e.g. waiting time priority (WTP[18]), Proportional Average Delay (PAD [21]) and the Hybrid Proportional Delay (HPD [18]). The difference between these schedulers is related to their speed of convergence under heavy and light load in the network, but it should be noted that all three schedulers use time dependent queueing delays to assign priorities to packets in different fashions.

Furthermore, Dovrolis et al in [20] introduced a method to provide an absolute guarantee to the end user through the proportional differentiated services by adding a dynamic class selection mechanism, where the application can increase/decrease its traffic class dynamically based on the QoS feedback reports from the receiver to satisfy its requirements.

In the rest of this paper, we will adopt the WTP algorithm which was studied in [22] with the name of Time Dependent Priority (TDP), and which was used by Dovrolis in [21] as an effective means to achieve the proportional delay differentiation in IP network. In WTP, the classifier adds $t_{arrival}$ to the header of each packet and forwards it to the corresponding queue according to its belonging priority class. The scheduler serves packets from queues in the FIFO manner by calculating the waiting time of each of head of line (HOL) packet (denoted by $w_p(t) = t - t_{arrival}$) in each queue and chooses the packet with the higher associated priority given by the following formula:

$$p_i(t) = \frac{w_p(t)}{\delta_i} = \frac{t - t_{arrival}}{\delta_i}$$

The scheduler selects packet with the largest priority value at time t from the HOL packets of all backlogged classes to be forwarded, according to the following formula:

$$ser\_p(t) = \arg_{i=1 \ldots N} \max(p_i(t))$$

Where N is the set of all backlogged classes. If two packets have the same priority value at time t then they will be transmitted in a random order, but the arrival process of traffic usually follows the Poisson distribution probability where the Pr(2 packets arrive at the same instant) is zero.

The different classes must have an equal waiting time priority at the same node in order to make the required proportionality between classes hold, e.g. transmitted packets of class i and j at time t_1 and t_2 must have:

$$\frac{w_i(t_1}{\delta_i} = \frac{w_j(t_2}{\delta_j} \quad \forall i, j \in \{1, 2, \ldots, N\}$$

While this mechanism is suitable for wired networks, it is still desirable to use this model in the wireless domain. Due to the fact that WTP is a centralized scheduling scheme, it needs to know the waiting times of all packets before deciding which one to transmit at a time. This is trivial in IP network, where all packets waiting to be scheduled originate from the same routers.

Due to shared medium and distributed access mechanism in ad hoc networks (CSMA/CA used in IEEE 802.11 [23]), WTP can not achieve proportionality between classes at the same node, because of the contention based access and the additional random probabilistic waiting time. In contrast to IP networks where the link is controlled by one router, frames underlying different classes at the MAC layer in ad hoc, wait for an additional random time before transmission (discrete uniform random variable), and thus render PDS inefficient with these kinds of networks. This additional random time may cause priority reversal at transmission time, because at this instance, the frame may no longer be the corresponding one to the paquet with the highest priority. For clarification, if packet $Packet_n$ received at MAC layer at time t_n, it will not be transmitted immediately but after a uniformly distributed random time. At transmission time t_{tx}, this packet may no longer have the largest waiting time priority $p_i(t)$ as shown in figure 1.

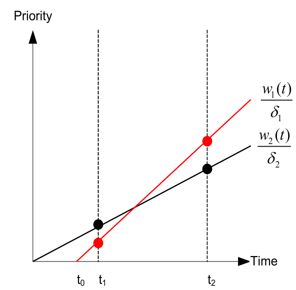


Fig. 1. Priority reversal

Most of the existing distributed QoS schemes for ad hoc provide service differentiation by proposing a new medium access mechanism, e.g. the use of priority based channel access [24], [25], or distributed fair scheduling [26], or linear mapping from network to MAC access priority [27],etc. However, even if some works have shown that it can achieve a relative service differentiation, where higher priority classes have better performance than lower priority classes, they did not provide a means to adjust the degree of differentiation between service classes, nor a formal proof for the differentiation result.

Recently, the IEEE Task Group proposes the Enhanced Distributed Coordination Function (EDCF) in IEEE 802.11e [28], which enhances IEEE 802.11 DCF with the introduction of different traffic classes by the use of distinct Arbitration Inter Frame Spaces ($AIFS_i$) and contention window (CW_i) sizes for different classes. We will exploit this access mechanism with the Markov analysis given in [29] to provide differentiation between classes.

III. Specification of the Proposed Model

Our objective is to extend PDS model in order to provide end-to-end proportional delay differentiation between classes in ad hoc networks. Our proposed model is constructed by composition of many mechanisms: priority adaptor, proportional differentiation scheduling mechanism, enhanced distributed prioritized medium access EDCF of IEEE 802.11e and delay estimator component as shown in figure 2.

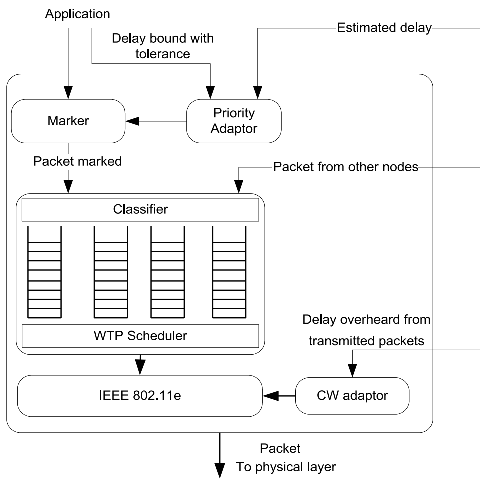


Fig. 2. QoS assurance mechanisms at each node

This mechanism works as follows: each application specify its QoS requirements parameters (maximum supported end-to-end delay, tolerated jitter, minimum throughput, etc.) to the priority adaptor component. This last is used to specify the appropriate priority dynamically in order to meet the QoS application requirements for its flow. The WTP scheduling mechanism is used at the network layer to provide differentiation between tagged packets in the same manner as in IP networks and to forward packets to the IEEE 802.11e layer. At link layer, a modification to the initialization fashion of control parameters in the MAC layer are proposed to provide proportionality between different access categories and packet delay fairness in each class at every node. This modification will be achieved through the use of congestion window adaptor component. In the next sub-section, we give a detailed specification of the tasks of each of these components.

A. IEEE 802.11e and contention window adaptor

In ad hoc networks, nodes access the medium with a decentralized scheduling scheme such as the distributed coordination function (DCF [23]) of IEEE 802.11. It is based on Carrier Sense Multiple Access Collision Avoidance mechanism (CSMA/CA) with binary exponential backoff algorithm when collision occurs.

EDCF in IEEE 802.11e is an extension to DCF mechanism for supporting QoS differentiation. The EDCF basic access method [28] is shortly summarized as follows: each packet from the higher layer arrives at the MAC layer with a specific priority value. A 802.11e station implement four access categories (ACs), where each packet arriving at the MAC layer with a pre-defined priority (traffic categories) is mapped into the corresponding AC. Basically, EDCF uses different Arbitration Interframe Spacing ($AIFS(AC_i)$), minimum Contention Window value ($CW_{\min}[AC_i]$) and maximum Contention Window value ($CW_{\max}[AC_i]$) for differentiation between packets belonging to the different ACs in contention phase to access the channel, instead of single $DIFS$, $CW_{\min}$, and $CW_{\max}$ values as in 802.11 DCF. These parameters will be exploited to provide proportional differentiation between ACs and consistent equal delay for each classes at every node.

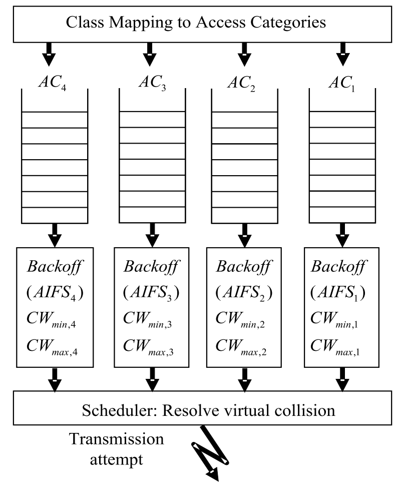


Fig. 3. Four ACs for EDCF.

Figure 3 shows the 802.11e MAC with four transmission queues in a station, where each queue behaves as a single enhanced DCF contending entity, i.e., an AC_i with its own $AIFS(AC_i)$ and Backoff Timer ($BT(AC_i)$). After sensing the channel idle for time period equal to $AIFS(AC_i)$ by an AC_i, it generates a random backoff value before transmitting. The backoff time counter is decremented $BT(AC_i) = BT_{old}(AC_i) - 1$ as long as the channel is sensed idle for unit time ($aSlotTime$). The value of $BT(AC_i)$ is freezed when a transmission is detected on the channel, then reactivated when the channel is sensed idle again for more than $AIFS(AC_i)$. The AC_i transmits when the backoff time $BT(AC_i)$ reaches zero. Moreover, the backoff timer is generated as $BT_i = random(0, CW_{i,j}) \times aSlotTime$, where $random()$ is a generator of random uniformly distributed from $[0, CW_{i,j} - 1]$ interval, and $aSlotTime$ is a very small time period (9μs) and $CW_{i,0} = CW_{\min}(AC_i)$ at the first attempt. After each unsuccessful transmission, $CW_{i,j}$ is increased exponentially by a factor 2 up to a maximum value $CW_{\max}(AC_i)$ as shows equation 2.

$$CW_{i,j} = \left\{ \begin{array}{ll} 2^j CW_{i,0} & 0 \leq j \leq m \\ 2^m CW_{i,0} & m \leq j \leq R \end{array} \right\} \qquad (2)$$

R is the retransmission limit at the MAC layer and it is equal to 7 in both DCF and EDCF. $CW_{i,j}$ denotes contention window of class i after a number of unsuccessful transmission j. After, a successful transmission, $CW_{i,j}$ will be reset to $CW_{i,0}$. When more than one AC_i within a station have their $BT(AC_i)$ expire at the same time, the collision is handled in a virtual manner. The highest priority packet among the colliding packets is chosen and transmitted, and the other queue performs the backoff mechanism while increasing $CW(AC_i)$ values.

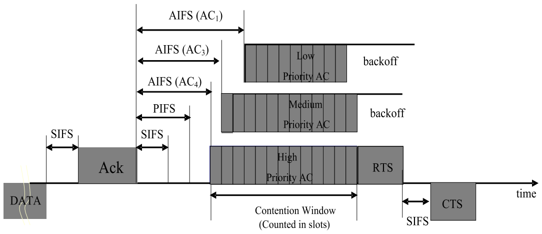


Fig. 4. EDCF channel access mechanism.

The basic access medium in EDCF is shown in Figure 4. This figure shows the timing diagram of the EDCF channel access. Basically, the smaller $AIFS(AC_i)$, $CW_{\min}[AC_i]$, and $CW_{\max}[AC_i]$, the shorter the channel average access delay for the corresponding priority, and hence the more capacity this priority obtains. However, the probability of collisions increases when operating with smaller $CW_{\min}[AC_i]$.

As Malli et al. in [30] show through simulations, EDCF performs poorly when the medium is highly loaded. This is due to the high collision rate and wasted idle slots caused by backoff in each contention cycle. The best medium distribution can be obtained only when EDCF supports a perfect scheduling algorithm among all the queues even in those at different nodes. That involves a complete synchronization and is difficult to realize in this kind of networks.

IEEE 802.11e achieve distributed priority scheduling by moderating the contention behavior in enhanced distributed coordination function (EDCF) [28] where better delays are provided using higher medium access priorities. The service differentiation is qualitative and does provide neither any specific delay assurance nor proportionality between classes. Consequently, we study the influence of the control parameters ($CW_{\min}[AC_i]$, $CW_{\max}[AC_i]$, $AIFS(AC_i)$, etc.) that affect PDS. As a result, we knew that $CW_{\min}[AC_i]$ is the most effective parameters in providing a relative differentiation between classes. Therefore, we should search the relationship between delay constraint and $CW_{\min}[AC_i]$ for providing a consistent differentiation.

Bianchi in [29] presents a Markov chain model for the analysis of the IEEE 802.11 saturation throughput. By analyzing the chain in the way proposed by Chatzimisios et al in [31], we get all required value for average transmitting delay experienced by contending nodes:

$$D_i = \sum_{j=1}^{R} P_{i,j} T_{i,j}$$

where $P_{i,j}$ is the probability that a node i transmits the frame at the $j-th$ backoff stage, and $T_{i,j}$ is the average delay. These parameters are given by Liqiang et al in [32], with:

$$P_{i,j} = (1-P_i)(P_i)^j \qquad 0 \le j \le R.$$

and

$$T_{i,j} = AIFS_i + \frac{S_i}{2} . \sum_{l=0}^{j} (W_{i,l}-1) + j.P_i.T_{i,c} \qquad 0 \le j \le R.$$

with S_i is the normalized throughput, P_i is the collision probability and $T_{i,c}$ is the collision time. Interested reader must refer to [31] for a detailed explication. Consequently, the ratio of the average delay of adjacent classes can be written as:

$$\frac{D_{i+1}}{D_i} = \frac{\sum_{j=0}^{R=7} P_{i+1,j} T_{i+1,j}}{\sum_{j=0}^{R=7} P_{i,j} T_{i,j}} =$$

$$\frac{\sum_{j=0}^{R=7} \Big(AIFS_{i+1} + \frac{S_{i+1}}{2} \sum_{b=0}^{j} (CW_{i+1,b}-1) + j \cdot p_{i+1} \cdot T_{i+1,c} \Big) \cdot P_{i+1,j}}{\sum_{j=0}^{R=7} \Big(AIFS_{i} + \frac{S_{i}}{2} \sum_{b=0}^{j} (CW_{i,b}-1) + j \cdot p_{i} \cdot T_{i,c} \Big) \cdot P_{i,j}}$$

As shows the previous formula, the control parameters that affect the delay experienced by a packet are: $AIFS_i$, R, $j.p_i.T_{i,c}$, $P_{i,j}$ and the minimum contention window size through $CW_{i,\min}$. Yang and Kravets in [33], minimum contention window sizes and throughput have been shown to have the following relationship, where we can conclude the equivalence through the relation between throughput (TH_i) and transmission delays (D_i):

$$\frac{TH_{i+1}}{TH_i} \approx \frac{\frac{L_{i+1}}{CW_{i+1,\min}}}{\frac{L_i}{CW_{i,\min}}} \Leftrightarrow \frac{D_{i+1}}{D_i} \approx \frac{CW_{i+1,\min}}{CW_{i,\min}} \qquad (3)$$

This can be explained by the fact that p_i is the same for all the classes, $AIFS[i]$ and $j.p_i.T_{i,c}$ are smaller than the other term. Therefore, delay proportionality can hold between classes at the same nodes as in end-to-end, because the packet's end-to-end delay equals the sum of all per-hop delays along the path. On the other hand, it is unfair to penalize nodes relaying more traffic than others. We turn our attention to provide a fair delay for each class at different nodes along the path. We want to provide an extension to equation 1 that must hold locally, to make it hold along every node in the networks, according to:

$$\frac{d_i^p(t,t+\tau)}{d_j^q(t,t+\tau)} = \frac{\delta_i}{\delta_j}, \quad \forall\, i \neq j \; and \, \forall\, p \neq q \qquad (4)$$

The superscripts p and q represent the *id* of two nodes along the path. An equal time delay between contending nodes can be achieved through a dynamic adjustment of minimum contention window as follows:

$$CW_i^p(t_k) = CW_i^p(t_{k-1}) \times (1 + \eta \frac{\bar{d}_N^p(t_{k-1}) - \bar{d}_i^p(t_{k-1})}{\bar{d}_N^p(t_{k-1})}) \qquad (5)$$

with $\overline{w_i^p}(t) = \frac{t_{tx}-t_{arrival}}{\delta_i} = \frac{W^p_{packet}}{\delta_i}$ is the normalized delay experienced by a packet at node p, and $\overline{d_i^p}(t_k)$ is the average of this normalized delay calculated as follows:

$$\overline{d_i^p}(t_k) = \alpha\overline{w_i^p}(t_k) + (1-\alpha)\overline{d_i^p}(t_{k-1})$$

$$\overline{d_N^p}(t_k) = \delta\overline{d_i^p}(t_k) + \beta\overline{d_N^q}(t_k) + (1-\delta-\beta)\overline{d_N^p}(t_k)$$

$\bar{d}_N^p(t_k)$ denotes the estimated normalized delay of the network at node p and η is a small positive constant. Each node must estimate its average waiting time $\overline{d_i^p}(t)$ after the transmission of each packet using the RTT average formula, and the average waiting time of the network $\overline{d_N^p}(t)$ after overhearing of a packet transmitted in its contending zone. Clearly, RTS/CTS/DATA/ACK frames can piggyback the waiting time of each packet and the average network estimation delay for two reception zone away along the path. This information in the packet header may be used to estimate the average delay experienced by a packet in the previous hop of the network, and at the node forwarding other flows in the same reception zone. The basic idea is to equalize the average normalized delay between nodes along the path such that equation 4 is satisfied. Therefore, the contention window adaptor must adjust the minimum contention window accordingly, by comparing the average delay of its transmitting packets with the networks average delay estimated from collected data. To provide proportionality at the same node, contention window adaptor updates the minimum contention window of only one predefined class according to equation 5 after a successful transmission, and for other classes according to equation 3.

B. Network layer and waiting time priority scheduler

The classifier handles the received packets by forwarding them to the appropriate waiting queue, where they wait before transmission to the MAC layer. The WTP scheduling is used to provide differentiation at the network layer in the same manner as in IP networks, where packets are treated in a proportional manner. WTP selects the packet with the longest normalized waiting time and sends it to the MAC layer.

C. The Priority adaptor

The end-to-end delay (throughput) sensitive applications request a bounded maximum delay (minimum throughput) with a jitter bound (tolerance bound). At the source node, this mechanism is responsible for determining the suitable class for each traffic flow. It begins by tagging packets with the lowest priority and compares received QoS report feedbacks with the required one. If QoS parameters are not satisfied, it increments the priority by one until the perceived QoS is satisfied or stays in the same class if it reaches the maximum priority level N. The priority adaptor may select directly the adequate priority if there is available information from existing flows. We do not claim to provide a hard guarantee with this priority adaptor, which tries to meet required bounds without providing any guarantee if the network is not able to deliver requirements of the application. This mechanism begin with priority $C_i^0 = 0$ and increases the priority of the flow periodically after T time if the received feedback report is inadequate with applications requirements. This mechanism works as follow:

$$\begin{cases} C_i^0 = 1 \\ C_i^{k(T+1)} = C_i^{kT} + 1 & if\ (C_i^{kT} < N \wedge QoS_{par} \notin SAT) \\ C_i^{k(T+1)} = C_i^{kT} & if\ (C_i^{kT} = N \wedge QoS_{par} \notin SAT) \\ C_i^{k(T+1)} = C_i^{kT} - 1 & if\ (C_i^{kT} > 1 \wedge QoS_{par} \in SAT) \\ C_i^{k(T+1)} = C_i^{kT} & if\ (C_i^{kT} = 1 \wedge QoS_{par} \in SAT) \end{cases}$$

Where i is a flow indicator and SAT is the satification set of QoS parameters.

IV. Performance Evaluation

In this section we study the performance of the proposed scheme using simulations performed in the $QNAP$. Queueing model is used due to its flexibility in adding time to header of each packets and its offred facility in accessing HOL packets information from other queues. The formal specification of each component in our scheme has been described through the use of algebraic operators of architectural description langage $AEMILA$ [34], before the description in $QNAP$.

We have chosen a a small grid topology of (3×3) (figure 5), with linear mobility in the four directions for all nodes, except A and C supposed fixe. The destination and the sources from where the data have to be sent are randomly generated in addition to that from A to C. When node B fails for an exponential delay used to simulate mobility and link break when node moves out, existing traffic from A to C will travel through AE along the path AEC to reach required destination.

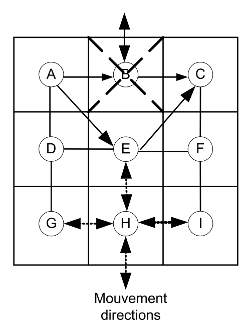


Fig. 5. Topology used in simulation.

We briefly describe the experimental setup and system configuration of our proposed model. Next, we present the results demonstrating the effectiveness of the proposed model in providing proportionality.

The robustness of our proposed EPDS scheme is tested using two different packet arrival profiles. Packets can belong to four different classes and usually the arrival process is described as a Poisson process. This process is the most widely popular traffic model because it takes into account the fluctuation of traffic. The time t between arrivals ($inter-arrival$) is exponentially distributed with rate λ:

$$Pr(t \leq T) = 1 - e^{-\lambda t}$$

and the number of arrivals in an interval of length t is then given by the Poisson probability:

$$Pr(n\ arrivals \in [0,t]) = \frac{(\lambda t)^n}{n!}e^{-\lambda t}$$

In contrast, recent studies and measurements show that realistic traffic follows heavy tailed distribution where the variance of data size is very large, even sometimes not finite and that can not be represented by Poisson distribution. Heavy tailed distributions are more convenient, e.g. Pareto distribution function given in equation 6 is an example of heavy tailed distribution. However, a robust model should not depend at distribution load assumptions for providing QoS.

$$Pr(t \leq T) = 1 - \frac{1}{t^{\kappa}} \quad (6)$$

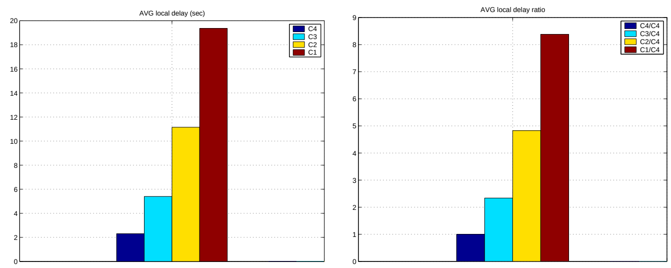
Therefore, we consider Pareto traffic arrivals for each class, where the packet arrival process follows the Pareto distribution with a shape parameter equals to $\kappa = 1.2$. All packets are constant length with 512bytes.

We first study the accuracy of EPDS model in providing differentiation between classes according to the pre-specified ratios at the same node and under the two arrivals pattern. We focus on scenarios of only four service classes at the network layer mapped directly to the 4 access categories used in IEEE 802.11e at MAC layer. All the parameters investigated in our model are given in table I. Results concerning the local average delay at a node are presented in figure 6, 7. It is obvious from these figures that average delay differentiation is mostly achieved simultaneously between different service classes according to their differentiation weight.

Parameters	*Value*
Number of classes at network layer	4
Number of classes at MAC layer	4
Differentiation parameters δ_i, $i \in \{1,2,3,4\}$	$\delta_1 = 1$, $\delta_2 = \frac{1}{2}$, $\delta_3 = \frac{1}{4}$, $\delta_4 = \frac{1}{8}$
MAC $CW_{i,\min}$, $i \in \{1,2,3,4\}$ for EDCF	[64, 32,16, 8]
MAC $CW_{i,\max}$, $i \in \{1,2,3,4\}$ for EDCF	1024
MacOverhead	28 Bytes
$aSlotTime$	$9\mu s$
$SIFS$	$16\mu s$
$DIFS = SIFS + 2 \times aSlotTime$	$34\mu s$
$AIFS_4$	$DIFS$
$AIFS_i = AIFS_{i+1} + aSlotTime$. $AIFS[4]$, $AIFS[3]$, $AIFS[2]$, $AIFS[1]$	$34\mu s$, $43\mu s$, $52\mu s$, $61\mu s$
Average weights α, δ, β	$\alpha = 0.9$, $\delta = 0.1$, $\beta = 0.1$
Per-class queue size (packets)	512bytes
Propagation delay	$1\mu s$
Delay jitter tolerance ε	20% of application delay

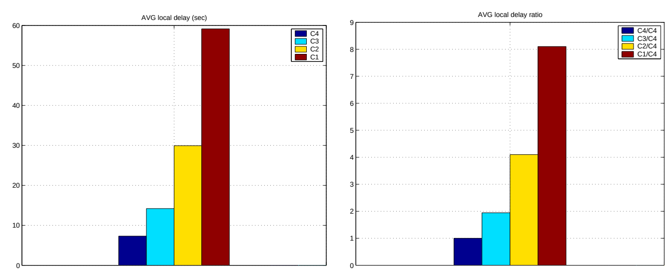
TABLE I
SIMULATION PARAMETERS.

User mobility leads to network topology changes after link breaking and thereby rerouting of all forwarded flows along the old path. When this occurs, traffic distribution changes significantly at other nodes in the same reception zone, and a transient perturbation of the proportionality ratio will occur and thus will result in short timescale violation of proportionality. This perturbation will not appear in the average and therefore a transient study is necessary to detect the influence of mobility at performance degradation. Figures 8, 9 show a non significant perturbation at a local node where proportionality between classes nearly continues to hold in the first 300sec of simulation run. The end-to-end delays proportionality continu hold perfectly with respect to differentiated parameters of $1 : \frac{1}{2} : \frac{1}{4} : \frac{1}{8}$, where we observe that the end-to-end achieved waiting time ratios are significantly closer to the target ratios. Velocity of each mobile node was taken 1m/sec during simulation.



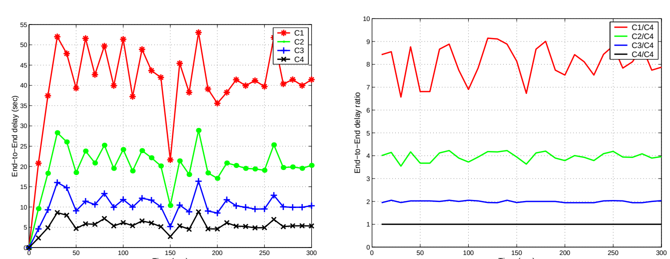
(a) Average local delay. (b) Average local delay ratio.

Fig. 6. Inter-arrival is exponentially distributed.



(a) Average local delay. (b) Average local delay ratio.

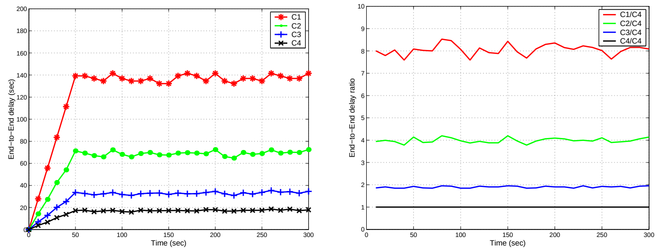
Fig. 7. Inter-arrival is Pareto distributed.



(a) End-to-End delay. (b) End-to-End delay ratio.

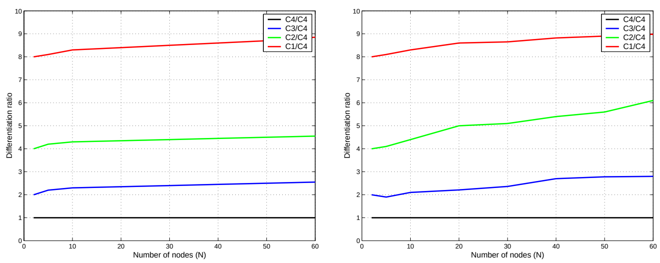
Fig. 8. End-to-End delay and delay ratio with Poisson distribution.

Then we extend the study to the impact of network size N at differentiation between classes. The variation curve is presented in figures 10(a) and 10(b) for Exponential and Pareto



(a) End-to-End delay. (b) End-to-End delay ratio.

Fig. 9. End-to-End delay and delay ratio with Pareto distribution.



(a) Exponential inter-arrival. (b) Pareto inter-arrival.

Fig. 10. Impact of network size.

inter-arrival pattern respectively. We observe that the achieved differentiation ratios are nearly equivalent to their assigned differentiation weights when the network size is small $(N \leq 10)$. In contrast, when the network size is large $(e.g. N \geq 40)$, our scheme tries to maintains a differentiation index close to the target, but it suffers from the number of collision that grows exponentially with the number of nodes (N).

In addition, our model results in a significant performance gain over EDCF that initializes its parameters in a static manner , regardless of channel condition. The gain appears in terms of enhanced throughput, reduced access delai and reduced collision probability even under a large size networks.

V. CONCLUSION

In this paper, we have investigated the problem of delivering high priority packets without over-compromising low priority classes by controlling the quality spacing between different classes. We study the problem of providing proportional delay differentiation by the use of EDCF and PDS. We showed the impact of tuning selected parameters of EDCF mechanism of IEEE 802.11e to provide and maintain service differentiation in the channel.

We investigate the impacts of different arrival pattern rules and show that our proposed scheme has the ability to softly re-adjust bandwidth among different types of traffic, in contrast to current QoS differentiation mechanisms that depend on specific assumptions of the distribution of traffic inter-arrival pattern. Our scheme makes the performance of network adaptively configurable by themselves which will minimize the impact of mobility at performance parameters.

From the performance point of view, we can also observe that our model scheme is an efficient way in providing differentiation between classes in predictable and controllable way. Moreover, our scheme is easy to implement and work in a completely distributed fashion. Finally, it is also possible to incorporate any proportional scheduling mechanism other than Waiting Time Proirity (WTP), to provide better support for differentiated services in mobile ad hoc networks.

REFERENCES

[1] S. Corson and J. Macker, "Mobile ad hoc networking (manet): Routing protocol performance issues and evaluation considerations," RFC 2501, January 1999.

[2] J. Broch, D. B. Johnson, and D. A. Maltz, "The dynamic source routing protocol for mobile ad hoc networks," IETF draft, 16 April 2003.

[3] C. E. Perkins, E. M. Belding-Royer, and S. Das, "Ad hoc on demand distance vector (aodv) routing," RFC 3561, July 2003.

[4] C. E. Perkins and E. M. Belding-Royer, "Quality of service for ad hoc on-demand distance vector," Internet Draft, 14 October 2003.

[5] J. L. R. Leung, E. Poon, A. lot Charles Chan, and B. Li, "Mp-dsr: A qos aware multi path dynamic source routing protocol for wireless ad hoc networks," in *the Proceedings of the 26th Annual IEEE Conference on Local Computer Networks (LCN'01)*, Tampa, Florida, 14–16 November 2001.

[6] J. Broch, D. A. Maltz, D. B. Johnson, Y. Hu, and J. Jetcheva, "A performance comparison of multi-hop wireless ad hoc network routing protocols," in *the Proceedings of MobiCom*. ACM Press, 1998, pp. 85–97.

[7] V. D. Park and M. S. Corson, "A highly adaptive distributed routing algorithm for mobile wireless networks," in *the Proceeding of IEEE Conference on computer communications (INFOCOM97)*, Kobe, Japan, April 1997, pp. 1405–1413.

[8] J. Xue, P. Stuedi, and G. Alonso, "Asap: An adaptive qos protocol for mobile ad hoc networks," in *the Proceeding of 14th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (IEEE PIMRC 2003)*, Beijing, china, 7-10 September 2003.

[9] R. Sivakumar, P. S. Inha, and V. Bharghavan, "Cedar: A core-extraction distributed ad hoc routing algorithm," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 8, pp. 1454–1465, August 1999.

[10] R. raden, D. Clark, and S. Shenker, "Integrated services in the internet architecture: An overview," RFC 1633,USC/Information Sciences Institute, MIT, Xerox PARC, June 1994.

[11] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An architecture for differentiated services," RFC 2475, December 1998.

[12] C. Bonnet and N. Nikaein, "A glance at quality of service models for mobile ad hoc networks," in *The proceedings of 16th Congrs DNAC (De Nouvelles Architectures pour les Communications)*, Paris, France, 2-4 December 2002.

[13] I. Chlamtac, M. Conti, and J. N. Liu, "Mobile ad hoc networking: imperatives and challenges," *Elsevier Ad Hoc Networks Journal*, vol. 1, no. 1, pp. 13–64, July 2003.

[14] S. Lee and A. Campbell, "Insignia: In-band signaling support for qos in mobile ad hoc networks," in *the Proceedings of Mobile Multimedia Communications (MoMuC)*, October 1998.

[15] G. S. A. S-B. Lee and A. T. Campbell, "Improving udp and tcp performance in mobile ad hoc networks with insigna," *IEEE Communications Magazine*, pp. 156–165, June 2001.

[16] A. L. Hannan Xiao, Winston K.G. Seah and K. C. Chua, "A flexible quality of service model for mobile ad hoc networks," in *IEEE Vehicular Technology Conference*, Tokyo, Japan, May 2000, pp. 445–449.

[17] A. V. Gahng-Seop Ahn, Andrew T. Campbell and L.-H. Sun, "Swan: Service differentiation in stateless wireless ad hoc networks," in *the Proceedings of IEEE INFOCOM 2002*, June 2002.

[18] P. R. Constantinos Dovrolis, Dimitrios Stiliadis, "Proportional differentiated services: Delay differentiation and packet scheduling," *IEEE/ACM Transactions on Networking (TON)*, vol. 10, no. 1, pp. 12–26, February 2002.

[19] M. K. H. Leung, J. C. S. Lui, and D. K. Y. Yau, "Adaptive proportional delay differentiated services: characterization and performance evaluation," *IEEE/ACM Transactions on Networking*, vol. 9, no. 6, pp. 801–817, December 2001.

[20] C. Dovrolis, "Dynamic class selection: From relative differentiation to absolute qos," in *ICNP '01: Proceedings of the Ninth International Conference on Network Protocols (ICNP'01)*. Washington, DC, USA: IEEE Computer Society, September 2001, pp. 120–128.

[21] P. R. Constantinos Dovrolis, Dimitrios Stiliadis, "Proportional differentiated services: Delay differentiation and packet scheduling," in *Proceedings of SIGCOMM*, Cambridge, Massachusetts, United States, 1999, p. 109120.

[22] L. Kleinrock, "A delay dependent queue discipline," *Naval Research Logistics Quarterly*, vol. 11, no. 4, pp. 329–341, December 1964.

[23] IEEE, "IEEE Std. 802.11: Wireless LAN Media Access Control (MAC) and Physical Layer (PHY) Specifications," 1999.

[24] X. Yang and N. H. Vaidya, "Priority scheduling in wireless ad hoc networks," in *MobiHoc '02: Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking & computing*. New York, NY, USA: ACM Press, 9 –11 June 2002, pp. 71–79.

[25] J. L. Sobrinho and A. S. Krishnakumar, "Real-time traffic over the ieee 802.11 medium access control layer," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 172–187, 1996.

[26] S. Gupta, N. H. Vaidya, and P. Bahl, "Distributed fair scheduling in a wireless lan," in *Sixth Annual International Conference on Mobile Computing and Networking*, Boston, August 2000.

[27] Y. Xue, K. Chen, and K. Nahrstedt, "Achieving proportional delay differentiation in wireless lan via cross-layer scheduling," *Journal of Wireless Communications and Mobile Computing, Special Issue on Emerging WLAN Technologies and Applications*, vol. 4, pp. 849–866, 2004.

[28] S. Mangold, S. Choi, P. May, O. Klein, G. Hiertz, and L. Stibor, "Ieee 802.11e wireless lan for quality of service," in *European Wireless*, Florence Italy, 26-28 February 2002, pp. 32–39.

[29] G. Bianchi, "Performance analysis of the ieee 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 535–547, 3 March 2000.

[30] M. Malli, Q. Ni, T. Turletti, and C. Barakat, "Adaptive fair channel allocation for qos enhancement," in *the proceedings of IEEE International Conference on Communications (IEEE ICC04)*, Paris, France, July 2004.

[31] P. Chatzimisios, V. Vitsas, and A. C. Boucouvalas, "Throughput and delay analysis of ieee 802.11 protocol," in *the Proceedings of the 5th IEEE International Workshop on Networked Appliances*, L. J. M. University, Ed., no. ISBN:0-7803-7686-2, UK, 30–31 October 2002, pp. 168–174.

[32] L. Zhao and C. Fan, "Enhancement of qos differentiation over ieee 802.11," *WLAN IEEE Communications Letters*, vol. 8, no. 8, pp. 494–496, August 2004.

[33] Y. Yang and R. Kravets, "Distributed qos guarantees for realtime traffic in ad hoc networks," in *the Proceedings of the IEEE International Conference on Sensor and Ad Hoc Communications and Networks (SECON-2004)*, October 2004, pp. 118–127.

[34] A. Aldini and M. Bernardo, "On the usability of process algebra: an architectural view," *Theoretical Computer Science*, vol. 335, no. 2-3, pp. 281–329, 2005.