Rate Allocation and Buffer Management for Proportional Service Differentiation in Location-Aided Ad Hoc Networks

Sivapathalingham Sivavakeesar, and George Pavlou

Centre for Communication Systems Research University of Surrey Guildford, Surrey GU2 7XH {S.Sivavakeesar, G.Pavlou}@surrey.ac.uk

Abstract. Given that applications and services for evolving mobile ad hoc networks (MANETs) have diverse quality of service requirements in a similar fashion to fixed networks, this paper proposes a proportional service differentiation (PSD) model. This model is highly scalable and simple to adopt in MANETs because it does not require explicit admission control or maintenance of state information in any intermediate node. It relies instead on localized scheduling and buffer management to achieve a desired global objective. Motivated by this aspect of the PSD model, we propose to combine it with a location-based forwarding strategy as a way to facilitate cross-layer optimization. This association is performed with a view to improve end-to-end service differentiation, although no other explicit mechanisms are used to achieve end-to-end guarantees. This model takes also into consideration the time-varying nature of available bandwidth in MANETs, and tries to calculate it dynamically. Simulation results confirm the per-hop performance improvement.

1 Introduction

The emergence of diversified multimedia applications requires that mobile ad hoc networks (MANETs) are augmented with service differentiation mechanisms, so that certain applications/services and users can possibly benefit from better quality of service (QoS) [2][3][5]. Service differentiation enables categorization of traffic into a set of classes to which network nodes provide priority-based treatment. Although service differentiation can be absolute or relative in nature [16], relative differentiation is preferred in ad hoc networks given that random node mobility, bandwidth and energy constrained operation and the unpredictable behavior of radio channels require a cost-effective solution [5][6]. In addition, absolute differentiation requires sophisticated admission control and resource reservation mechanisms, which are difficult to achieve in highly dynamic ad hoc networks. A relative differentiation mechanism that supports a small number of service classes is simpler in terms of implementation, deployment and manageability [7]. However, relative service differentiation (RSD) can only provide weak guarantees, which do not always address the requirements of

applications [8]. In the case of RSD, there is no absolute guarantee that high-priority classes will perform better than lower-priority ones, and this varies depending on the load of each class. In order to tackle this problem, recent research studies have aimed to strengthen the service assurance provided by RSD without incurring much additional complexity. The proportional service differentiation (PSD) model is the result of such studies, and defines a service model with no admission control or an explicit resource reservation mechanism [7][8]. It supports a certain number of service classes relatively ordered in terms of loss rate or average queuing delay and the ratios of those QoS metrics between successive priority classes remain roughly constant, irrespective of network load [7], [8], [9], [10], [11], [12]. The PSD model exhibits controllability and predictability, which make it different from RSD [7].

Service differentiation at MAC-level in terms of prioritized access to the common wireless medium using the DCF-based operation of IEEE 802.11 a/b has been proposed in the literature [3], [4]. Although our work does not address such MAC-level differentiation, it considers the fact that the wireless channel in MANETs is a shared-access medium in which the available bandwidth varies with the number of hosts contending for access [15]. Hence, our work adopts an approach of determining the effective bandwidth dynamically, and this is important for arriving at a service rate allocation among the different service classes. At the network level, we consider only per-hop service differentiation with the understanding that the adoption of such localized behavior at each node will result in end-to-end service differentiation. The latter is facilitated by our forwarder-node selection strategy that identifies and routes packets along non-overloaded mobile nodes (MNs). With the adoption of effective bandwidth calculation mechanism and our forwarder-node selection algorithm, the proposed model attempts to alleviate two major challenges in mobile ad hoc networks; i) fluctuating bandwidth at each node, and ii) topology changes due to mobility.

The rest of this paper is organized as follows. After providing the description of the problem statement in the next subsection, previous work on MANET quality of service and proportional service differentiation is reviewed in section 2. Section 3 presents our model and section 4 evaluates it through simulations. Finally, section 5 concludes the paper and points to future work.

1.1 Problem Statement

The proposed model ensures proportional service differentiation over multiple QoS metrics (packet loss rate and queuing delay) among classes at a mobile ad hoc node (i.e. single-hop only). Hence, the problem is to develop scheduling (service rate allocation) and buffering management policies that each node can use to service multiple competing classes in order to satisfy the QoS and system constraints such as maximum available buffer size and time-varying link capacity at each node. The service rate allocation can thus be viewed as an optimization problem performed in a distributed fashion and subject to the above QoS and system constraints.

2 Previous Work

2.1 Service Differentiation in MANETs

Service differentiation in MANETs has originally focused on MAC design, especially tailored to the IEEE 802.11 DCF-based operation [2], [3]. Service differentiation is achieved by setting different values for the lower and upper bounds of contention windows for different service classes. There is however, no explicit guarantee of the level of service differentiation. Due to this reason, although such MAC-level differentiation is beneficial, strong per-hop class differentiation is possible through the adoption of PSD model as described in this paper. More recently, network-level service differentiation was addressed in MANETs, with a stateless QoS framework that uses rate control for best-effort traffic and source-based admission control for realtime traffic [4]. This model is called SWAN (service differentiation in stateless wireless ad hoc networks) and any source can admit its own flow based on sending probing-requests towards a destination. Although this approach claims to be stateless, intermediate nodes may be required to remember whether the flows that traverse them are new or old in order to regulate traffic [4]. In addition, source-based admission control using probing-packets is unrealistic in a dynamic environment like MANETs, as conditions and network topology tend to change fairly frequently. Bandwidth calculations do not take best-effort traffic into consideration, and hence may lead to a false estimation of the available bandwidth.

Very recently, relative bandwidth service differentiation was proposed in [5]. The service profile for a traffic flow is defined as a relative target rate, which is a fraction of the effective link capacity of nodes. This flow-based (as opposed to class-based) approach is ambiguous and unrealistic as service profiles are arbitrarily assigned to flows. It does not show how a target rate for a particular flow is arrived at in any node and passed to other nodes along a specific route. This approach has an important drawback, which becomes clear mainly in shorter timescales, unless these target rates are adjusted dynamically based on performance measurements. The reason for this behavior is that the service quality in each class depends on the short-term relationship between the allocated services to a class and the arriving load in that class [7]. As a result, a higher class can often provide worse QoS than lower classes, invalidating the main premise of relative service differentiation.

Another work on proportional differentiation considered mainly delay differentiation in a WLAN environment [6]. However, our approach attempts for the first time to support both proportional loss and delay differentiation in a dynamic environment such as an ad hoc network.

2.2 Related Work on Proportional Service Differentiation

The proportional service differentiation model was first introduced as a per-hopbehavior (PHB) in the context of wireline differentiated services networks [7]. Two key modules are required to realize PSD: packet scheduling and buffer management. The common approach is to use scheduling algorithms for delay differentiation and use buffer management for loss differentiation. Two scheduling approaches, the proportional queue control mechanism (PQCM) and backlog-proportional rate (BPR) dynamically adjust class service rate allocations to meet QoS requirements [9]. A waiting-time priority (WTP) scheduler operates on dynamic time-dependent priorities, while there exist a number of WTP variations in the literature. Buffer management functionality can be split into the backlog controller and the dropper. The backlog controller specifies when packets need to be dropped, while the dropper actually drops them. Random early detection (RED) is a typical example of a backlog controller, whereas drop-tail is a widely used dropping mechanism [9].

3 The Proposed System Model

3.1 Model Description

The proportional delay and loss differentiation model is implemented with the use of a packet-forwarding engine. This consists of buffer and scheduling units [9]. The packet buffer is logically organized into Q queues, one for each class (where Q is total number of classes of service). These Q queues share the physical link bandwidth and the buffer space. Scheduling is assumed to be work-conserving, and the scheduler dynamically allocates bandwidth to each of the Q service classes in order to achieve delay differentiation. When the packet buffer is full, the buffer management unit will select and drop certain number of packets from the tail of a particular class in order to meet the proportional loss rate constrains. A first-come-first-served policy is used to transmit traffic from the same class. Although our scheduler inherits the idea of the time-dependent priority concept from WTP, it has been sufficiently augmented with predictions – and hence, our scheduler is called predicted delay proportional (PDP) scheduler. The PDP scheduler is based on a fluid traffic model.

Given that there exist Q service classes which are ordered, so that class-i is better than class-j for $i \neq j$, $1 \le i < Q$ and $i < j \le Q$. Assume that $\overline{d}_i(t,t+\tau)$ is the average queuing delay of packets of class-i in the interval (t, t+ τ), where $\tau > 0$ is the monitoring timescale, then for all classes i and j, the following needs to be satisfied per-hop independently of available bandwidth and class loads:

$$\frac{\overline{d}_{i}(t,t+\tau)}{\overline{d}_{i}(t,t+\tau)} = \frac{\zeta_{i}}{\zeta_{i}} \tag{1} \qquad \frac{\overline{l}_{i}(t,t+\tau)}{\overline{l}_{i}(t,t+\tau)} = \frac{\sigma_{i}}{\sigma_{i}} \tag{2}$$

 ζ_i are delay differentiation parameters (DDPs), being ordered as $\zeta_1 < \zeta_2 < ... < \zeta_Q$. In case of loss rate differentiation, for the fraction of packets of a specific class that were backlogged at time t or arrived during the interval (t, t+ τ) and were dropped in this same time interval, the above proportional loss rate differentiation per single hop as given by equation (2) should hold. In this case, σ_i are the loss differentiation parameters (LDPs), being ordered as $\sigma_1 < \sigma_2 < ... < \sigma_Q$. In order to satisfy equations (1) and (2), instead of considering only the delay of a top packet of a queue, the PDP scheduler tries to predict the average delay of all the packets in the queue. In this process, the following are assumed during the monitoring interval [9]:

- i The service rate of a specific class is unchanged until the packet is dequeued.
- ii No packet is dropped.
- iii The packet loss rate of a specific class is unchanged.

After predicting the delays of all packets in a specific class queue, the mean delay of that queue is determined. If the predicted delay of each class follows the required proportional delay differentiation with respect to other service classes, the service rate and the loss rate associated with each class are not altered. On the other hand, if the predicted delays of one or many classes do not satisfy a constraint, then either the service rate or the loss rate for each backlogged class need to be changed as explained below. If there are no buffer overflows, the predictions for delay violations are made only once for every Y packet arrivals. The selection of a proper value for Y represents a tradeoff between the runtime complexity and performance improvement with respect to satisfying the constraints. On the other hand, when there is a buffer overflow, packets need to be dropped while still maintaining the constraints. In our work, we consider per-hop proportional loss rate and queuing delay constraints among classes. In case any of these proportional constraints lead to an infeasible system, some constraints need to be relaxed in a specific precedence order until the system becomes feasible. For this purpose, system constraints have priority over proportional constraints [9]. Since the service rate allocation is viewed as an optimization problem, the objective function aims at i) minimizing the amount of traffic to be dropped, and ii) maintaining the current service rate allocation. The first objective ensures that traffic is dropped only if there is no alternative way to satisfy the constraints, while the second is to minimize fluctuations in the service rate allocation to each class. As mentioned in the problem statement, this work considers the maximum buffer size and bandwidth available at each node as system constraints as follows: $\sum_i B_i(t) \le B_M$ and $\sum_i \mu_i(t) = C_M$, where $B_i(t)$ backlog of class-i at a time

instance t, B_M is the buffer size of any node M, $\mu_i(t)$ is the service rate allocated to class-i at a time instance t, and $C_M(t)$ is the time-varying link capacity available at any node M. Since the maximum link capacity available to a node vary with time in MANETs, we adopt a method to determine it dynamically as will be explained in section 3.2. Since our model makes delay predictions while keeping the loss rate of a class-i ($1 \le i \le Q$) constant in the monitoring timescale, the optimization problem can be formulated as follows:

Find new service rates $\mu_i(t)$ and $\mu_j(t)$ for backlogged service classes *i* and *j*, where $i \neq j$, $1 \leq i < Q$ and $i < j \leq Q$, such that at least the three conditions as given by equations (1) and those imposed by system constraints are satisfied.

The proportional loss rate dropper is a simple dropper sharing the idea of WTP scheduler and has two objectives: i) try to minimize the number of packets being dropped and ii) when there needs to be a packet drop, pick a packet from a certain class in order to keep the loss rate proportional, while satisfying the other constraints. The concept of a weighted loss rate is used for comparison purposes in order to make the packet dropping decision. Whenever a packet tries to join an already full buffer, the packet dropper makes a decision as to which priority packet it should drop in order to keep the loss rate of each class. The values of these parameters are used here to calculate the weighted loss rate of each class. The values of these parameters are chosen such that $w'_1 \times l_1 = w'_2 \times l_2 = \cdots = w'_Q \times l_Q$ is satisfied, where $w'_i \times l_i$ (t) is the weighted loss rate is then dropped to keep the loss rate differentiation proportional.

3.2 Link Capacity Estimation

Since we assume the DCF-based operation of IEEE 802.11 as the underlying MAC, the link capacity available to any node needs to be determined dynamically [13]. This is because the available bandwidth varies depending on several factors, namely node-mobility, network topology, power constraints, and contention from other neighboring nodes. In the estimation process we assume that each node employs RTS and CTS frame transmission at the MAC-level in order to minimize the hidden and exposed terminal problems [15]. If time instances t_1 and t_5 and frame size (fs) of the data of Fig. 1 are known, then any node M can predict the bandwidth available using equation (3) [13]:

Predicted Bandwidth Availability (
$$C_{M}(t)$$
) = $\frac{FrameSize}{t_{5}-t_{1}} = \frac{fs}{\Delta t}$ (3)



Fig. 1. IEEE 802.11 unicast packet transmission sequence This link layer measurement mechanism captures the effect of contention on available bandwidth. If contention is high, t5 - t1 would increase and bandwidth available would decrease. This mechanism also captures the effect of fading and interference,

since if RTS or CTS get lost, they need to be re-transmitted. This increases the Δt of equation (3), and hence would result in lower bandwidth. It should be noted that the available bandwidth is measured using only successful link layer transmissions. In addition, this estimation process uses average throughput of past packets to estimate the current bandwidth, and it has been proved feasible and robust [13].

3.3 Traffic Regulation

Although our present work does not consider any mechanism for explicit end-to-end absolute service guarantees, our model tries to use localized per-hop and per-node information to improve end-to-end performance. For this purpose, it works in conjunction with a location-based forwarding mechanism [1]. With the proper forwarder-node selection, the per-hop information is efficiently utilized to minimize congestion, and hence enables traffic regulation.

Sophisticated admission control and traffic policing mechanisms are normally used to regulate traffic, and such mechanisms normally necessitate maintenance of state information [7]. Admission control and policing lead to better performance in the case of fixed IP networks, where routes taken by packets are not volatile. On the other hand, they may not bring in a tangible improvement in the case of dynamic MANETs, as routes taken by packets of the same flow may vary heavily with time. Traffic can be regulated in such situations in a proactive manner by selecting proper non-overloaded forwarding nodes, which is possible in a mesh-like network.

$$\Omega_{\rm MI} = \frac{LET_{MI} * BW_{Available_{I}}}{d_{MI} + d_{IB}}$$
(4)

Let N(M) be the neighbor set of node M, and M currently have a packet to be forwarded, d_{MI} be the distance from node M to any of its one-hop neighbors I (I \in N(M)), d_{IB} be the distance from any node I (I \in N(M)) to the packet's destination B, LET_{MI} be link expiration time of M with respect to I (I \in N(M)), and BW_{Available} be the bandwidth available at node I (I \in N(M)). The criterion used in our forwarder selection algorithm is given by equation (4). The selection strategy considers the currently available bandwidth to a neighbor, link expiration time (LET) and relative locations of the node-pair under consideration [1]. Any neighbor I of node M that has the highest value for $\Omega_{\rm IM}$ of equation (4) can be chosen by our algorithm as a forwarding-node. This way of proper forwarder-node selection is essential in MANETs due to the following two reasons, i) node mobility may lead to a situation where the selected forwarder will soon move away from the sender so that the packet transmission will fail, and ii) the selected forwarder is so busy that the forwarded packets face long delays or get ultimately dropped. Hence, our forwarder selection strategy involves relative mobility prediction and dynamic bandwidth estimation as described below in order to minimize these undesirable aspects. If motion parameters of two neighbor nodes (e.g. velocity, radio propagation range) are known, there is a way to determine the duration of time these two nodes will remain connected [14]. The predicted time is the link expiration time (LET) between two nodes, and this is used in equation (4).

The bandwidth available to a node can be estimated using equation (4) only if t_1 and t_5 of Fig. 1 are known. In our strategy any node should be able to calculate the bandwidth available to any one of its neighbors by listening to the transmission initiated by the latter. However, it is difficult for any node to determine the exact time (t_1) at which a packet becomes ready for transmission in its neighbor. On the other hand, any node can be aware of time instances at which control frames associated to a particular data frame are initiated by one-hop neighbors by listening to the medium promiscuously. This MAC-level listening is facilitated in the DCF-based operation of IEEE 802.11, as virtual carrier sensing with the use of network allocation vector (NAV) is necessary for the correct DCF-based operation [13], [15]. If the time instance at which the RTS frame is initiated is known, then we can decide a value for t_1 empirically based on node-density and recent traffic characteristics. The frame size of the neighbor node can be statistically determined by taking a time average of the last k-number of packets it generated. Any node can determine this by either analyzing the packets received from its neighbor or by listening to the latter's transmissions promiscuously. Also, under certain circumstances, it may be difficult for a listening node to determine the exact time at which the neighbor receives an ACK. Hence, in such cases we have to determine a value for t_5 empirically by knowing the time at which the data frame transmission ends (t_4). The time instances t_1 and t_5 can be determined by the following two equations: $t_1 = t_2 - DIFS - \delta_1$ and $t_5 = t_4 + SIFS + t_{ACK}$ + δ_2 , where SIFS is short inter-frame space, DIFS is distributed inter-frame space, t_{ACK} is the transmission time for ACK frame, δ_1 is added to take care of an extra time involved to access the channel due to the binary exponential backoff mechanism of the DCF and δ_2 is added to take care of an extra time involved if data needs to be retransmitted due to collisions or channel errors [13], [15]. In order to find values for $\delta 1$ and $\delta 2$, we need to first analyze the binary exponential backoff mechanism of the DCF-based operation of the IEEE 802.11 [15]. For simplicity it is assumed that $\delta_1 =$ δ_{2} , and further we assume that each takes a value of an average backoff window in a saturated network condition. By determining a suitable value for the mean backoff window of an interested neighbor node, we can determine its bandwidth using equation (3) [13], [15].

4 Evaluation through Simulation

The GloMoSim simulation package [17] was used to evaluate how well proportional service differentiation is achieved in a per-hop manner using our model. Four service classes 1, 2, 3 and 4 were considered for this purpose, where class-1 had the highest priority. Proportional factors for packet loss rate and delay were chosen as 1:2:3:4 in our simulations. The routing protocol used was a location-based one, which had been augmented with our forwarder-node selection algorithm. Our implemented routing mechanism considered simple greedy forwarding only. Traffic was generated using random CBR connections having a payload size of 512 bytes. These CBR connections

tions were randomly generated so that at any moment the total number of sourcedestination pairs was kept constant – and each session lasted for a time-period that was uniformly distributed between 40 and 50 seconds. Each class contributed equally to the total traffic in the network (i.e. 25%). We considered two QoS-constraints, proportional loss rate and proportional delay. Metrics such as the average delay and loss performance were measured using a sliding window size of 0.5 second. A terrain of 600X 600 m² was considered with 80 nodes moving at a maximum speed of 10 ms⁻¹ and initiating 10 sessions at any instant. Fig. 3 (a) and Fig. 3 (b) depict per-hop absolute loss rates and absolute average queuing delays of all classes averaged over time intervals of length 0.5 s respectively, when the traffic is generated as a sinusoidal function of simulation time. As it can be seen from these two cases, our model maintains a consistent ratio between classes irrespective of the network load conditions.



Fig. 3. Absolute Per-hop Queuing Delays and Loss Rates as function of simulation time

5 Conclusions and Future Work

In this paper we presented a novel scheduling and buffer management mechanism to realize per-hop proportional loss and delay differentiation in ad hoc networks. In this context, we considered the time-varying nature of wireless link bandwidth available to any node in an ad hoc network and adopted an approach to estimate it. The proposed proportional service differentiation model works in conjunction with location-based forwarding mechanisms. Although the motivation behind this interaction is to facilitate end-to-end service guarantees, in this paper we did not include any explicit mechanism to achieve this objective. In the future, we will concentrate on how to incorporate end-to-end absolute service guarantees in our model through interactions between the PSD unit and the location-based forwarding engine. We will then evaluate how well our enhanced PSD model will guarantee both proportional and absolute service guarantees in an end-to-end manner.

References

- S. Sivavakeesar, and G. Pavlou : Cluster-based Location Services for Scalable Ad hoc Network Routing. Int. Conf. on Mobile and Wireless Communications Networks (MWCN' 2004), Paris, France, Oct. 2004, 433 – 448
- M.Barry, A.T.Campbell, and A.Veres, "Distributed Control Algorithms for Service Differentiation in Wireless Packet Networks", Proc. 20th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2001), vol. 1, Apr. 2001, pp. 582 – 590.
- I.Aad, and C.Castelluccia, "Differentiation Mechanisms for IEEE 802.11", Proc. 20th Annual Joint Conference of the IEEE Computer and Com. Societies (INFOCOM 2001), vol. 1, Apr. 2001, pp. 209 – 218.
- G-S.Ahn, A.T.Campbell, A.Veres and L-H.Sun: SWAN: Service Differentiation in Stateless Wireless Ad Hoc Networks. Proc. IEEE INFOCOM'2002, June 2002
- K.C.Chua, H.Xiao, and K.G.Seah : Relative Service Differentiation for Mobile Ad hoc Networks. *Proc.* IEEE Wireless Communications and Networking Conference (WCNC 2003), vol. 2, Mar. 2003, 1379 – 1384
- K-C.Wang, P.Ramanathan, "End-to-End Delay Assurance in Multihop Wireless Local Area Networks", Proc. Global Telecommunications Conference (GLOBECOM 2003), vol. 5, Dec. 2003, pp. 2962 – 2966.
- C.Dovrolis, and P.Ramanathan : A Case for Relative Differentiated Services and the Proportional Differentiation Model. IEEE Network, vol. 13, no. 5, Sep. 1999, 26 – 34
- C.Dovrolis, and P.Ramanathan : Proportional Differentiated Services, Part II: Loss Rate Differentiation and Packet Dropping. Proc. 8th Int'l Workshop on Quality of Service (IWQoS 2000), Jun. 2000, 53 – 61
- N.Christin, and J. Liebeherr : A QoS Architecture for Quantitative Service Differentiation. IEEE Communication Magazine, vol. 41, no. 6, Jun. 2003, 38 – 45
- 10. A.Striegel, and G.Manimaran : Packet Scheduling with Delay and Loss Differentiation. Elsevier Computer Communications, vol. 25, issue 1, Jan 2002, 21 31
- M.Hamdaoui, and P.Ramanathan: A Dynamic Priority Assignment Technique for Streams with (m, k)-Firm Deadlines. IEEE Transactions on Computers, vol. 44, no. 12, Dec. 1995, 1443 – 1451
- W.Lindsay, and P.Ramanathan : DBP-M: A Technique for Meeting End-to-End (m, k)-Firm Guarantee Requirements in Point-to-Point Networks. Proc. IEEE Conf. on Local Computers, 1997, 294 – 303
- S.H.Shah, K.Chen, and K.Nahrstedt : Dynamic Bandwidth Management for Single-hop Ad Hoc Wireless Networks. ACM/Kluwer Mobile Networks and Applications (MONET) Journal, Special Issue on Algorithmic Solutions for Wireless, Mobile, Ad Hoc and Sensor Networks, vol. 10, no. 1, 2005
- W.Su, S-J.Lee, and M.Gerla : Mobility Prediction and Routing in Ad Hoc Wireless Networks. Int. Journal of Network Management, vol. 11, no. 1, John Wiley & Sons, Jan.- Feb. 2001, 3 – 30
- O.Tickoo, and B.Sikdar : Queuing Analysis and Delay Mitigation in IEEE 802.11 Random Access MAC based Wireless Networks. Proc. IEEE INFOCOM'2004, vol. 2, Mar. 2004, 1404 – 1413
- S.Blake, D.Black, M.Carlson, E.Davies, Z.Wang, W.Weiss : An Architecture for Differentiated Services. IETF RFC 2475, Dec. 1998
- X. Zengu, R.Bagrodia, and M.Gerla : GloMoSim: A Library for Parallel Simulations of Large-scale Wireless Networks. Proc. of the 12th Workshop on Parallel and Distributed Simulations, May 1998