A Case for A Mobility Based Admission Control Policy

Shahram Ghandeharizadeh[1], Tooraj Helmi[1], Shyam Kapadia[1], Bhaskar Krishnamachari[1,2]
[1]Department of Computer Science
[2]Department of Electrical Engineering
University of Southern California
Los Angeles, CA 90089, USA

ABSTRACT

Ad hoc networks of wireless devices such as Car-to-Car Peer-to-Peer Networks (C2P2) are an emerging technology. Entertainment applications that manipulate continuous media, audio and video clips, push the limits of these networks due to their stringent requirements. These challenges are magnified by environmental characteristics such as dynamic wireless network connections, mobility, multi-hop nature of the network and the possible lack of a fixed infrastructure. In these networks, an intelligent admission control policy enhances Quality of Service (QoS) observed by the users of the system. This paper makes the case for a decentralized Mobility based ADmission Control (MAD) policy. We develop QoS utility models to quantify the performance of this policy with an environment that employs no-admission control. Obtained results show conclusively that MAD provides orders of magnitude improvement when compared with no admission control.

1. INTRODUCTION

During the past few years, automobile manufacturers have been marketing and selling vehicles equipped with entertainment systems. These systems typically consist of a DVD player, a fold-down screen, a video game console, and wireless headphones. In its present form, storage and content are tied together. This limits the number of available titles to those DVDs and CDs in the vehicle. We envision a separation of storage and content where content is staged on demand across the available storage for previewing. This would provide passengers access to a large repository of titles. In this vision, vehicles are equipped with car-to-car, peer-to-peer (C2P2) devices which form a mobile ad-hoc network (MANET). Each C2P2 is equipped with abundant amount of storage, a processor, and a wireless networking card.

The principle characteristic of continuous media such as video and audio is their high sustained bit rate requirement. If a system delivers a clip at a rate lower than its pre-specified rate without special precautions (e.g., pre-fetching), the user might observe frequent disruptions and delays with video and random noises with audio. These artifacts are collectively termed *hiccups*. A hiccup-free display is the ideal quality of service (QoS) provided to an end user. A second QoS criterion is the observed startup latency, defined as the delay observed from when a user references a clip to the onset of display.

All C2P2 devices may replicate popular titles, but they must collaborate in order to provide users with a large selection of titles. Each C2P2 may contribute a fraction of its storage in a peer-to-peer manner to be occupied by the system-assigned clips. These clips might be referenced by users of those mobile C2P2 devices that are network reachable (including the local user). A C2P2 must offer its passengers only those clips that it can download and play in a hiccup-free manner within a reasonable amount of time (e.g. by satisfying a maximum tolerable startup latency constraint).

Techniques that address these challenges are impacted by the characteristics of the MANET, placement and delivery scheduling of data, and admission control policies in support of a hiccup-free display. Consider each in turn. Mobility, a key MANET characteristic, is the primary challenge because it dictates the life-time of paths between a producer and a consumer of data. When the display of data is overlapped with its delivery, a path repair may incur delays that result in data starvation and hiccup. In addition, topology changes impact the availability of both data and bandwidth. The availability of data is also impacted by its placement across the nodes and its degree of replication. One may replicate data at the granularity of either a clip [1] or a block [2].

Finally, given the high data-rate requirements of continuous media displays and limited bandwidth resources, the system must be configured with an admission control policy to handle multiple simultaneous clip downloads. In particular, the admission control policy must strive to do a "good job" of rejecting those requests that are unlikely to be satisfied within a reasonable amount of time. Section 3 formalizes and quantifies this qualitative metric by providing utility models for QoS that consider the number of rejected requests, the number of admitted requests that are successful, and the number of admitted requests that fail to meet a specified startup latency constraint.

Compared to traditional admission control techniques, the C2P2 MANET environment offers the following unique challenges. First the network topology is dynamic. Second the admission control must be performed in a distributed manner as there is no central coordination point for the network. Further, it is hard to estimate the resources that will be available for a clip download (since the topology will change during the download). The key contribution of this paper is the design and analysis of a simple mobility based admission control policy using C2P2/MANET simulation studies. While admission control is well studied in wired networks [4, 8], to the best of our knowledge, besides our prior work with stationary wireless devices [2, 1], there is little if any prior work on this topic.

Recent studies addressing QoS in MANETs are as follows. A survey of QoS metrics in MANETs is provided by [10]. These studies either build on top of the existing routing protocols [11, 6] like DSR, AODV, etc., or integrate the QoS metrics into the routing protocol [5, 9, 12]. There have been studies on both unicast as well as multicast routing. Our primary contribution is to extend the conventional definition of admission control and introduce a simple mobility based admission control policy to study the number of rejected, satisfied and unsatisfied requests in a MANET (C2P2 network).

The rest of this paper is organized as follows. Section 2 describes the framework for the admission control problem and introduces the mobility based admission control policy. In Section 3, we show the performance observed in environments deployed with and without MAD. Due to space limitations a comparison of MAD with other admission control techniques is not presented here. Our conclusions are presented in Section 4.

2. ADMISSION CONTROL

Section 2.1 formalizes admission control as the process of admitting those requests that are able to download a fix-sized file within a specified duration of time. Section 2.2 develops the framework and describes our mobility based admission control policy.

2.1 Introduction and Problem Statement

While it is possible to overlap the delivery and display of continuous media clips using buffering, this strategy runs the risk of incurring hiccups if network resources becomes temporarily unavailable during delivery. In this work, we will focus our attention on the delivery of clips that will be played only after the entire clip is downloaded. We will assume that each request specifies the file size of the clip to be downloaded and a maximum download time (i.e. startup latency).

We formulate the admission control problem as follows: A request to display a clip X requires all blocks of X be downloaded to the requesting C2P2 (denoted as C2P2$_d$) within T_X time units. Note that T_X (the download time or the startup latency) can be significantly smaller than the playback time of the clip X. For example, an audio clip X with a display time of 12 minutes might specify T_X as 1 minute. This means that all blocks of this clip must be materialized in one minute in order for a request referencing X to be admitted to the system. Here, once a request is issued, the decision to either reject or admit the request must be made almost immediately after the request is issued (Instantaneous mode of operation).

Thus, assuming K requests are issued to a system, an admission control policy yield three classes of requests. First, a fixed number of admitted and rejected requests, termed A and R, respectively. Note $K = A + R$. Second, some of the admitted requests might fail, denoted as A_f, $A_f \leq A$. And, finally, some of the admitted requests are serviced successfully, denoted as A_g, where $A = A_g + A_f$. An admission control policy should strive to maximize both A and A_g (ideally, to a maximum of K). By maximizing A, a larger percentage of requests are processed by the C2P2 MANET. By maximizing A_g, a larger percentage of admitted requests are processed successfully. Section 3 presents QoS models that define a policy's utility as a weighted combination of these raw metrics.

2.2 Mobility-based ADmission Control (MAD)

We now propose a framework for solving the admission control problem in a mobile ad-hoc network. We focus on distributed, client-centric, admission control policies that enable the C2P2 client device to either admit or reject a request. For the rest of this paper, C2P2$_d$ is the device with a pending display request.

In order to make an intelligent admission decision, the admission control component of C2P2$_d$ must first obtain information about the state of the network and availability of resources. This information is then weighed against the specifications of the request in order to determine if the request can be admitted. Here, we present the results for the mobility based admission control policy.

In a mobile ad-hoc network such as C2P2, knowledge of the current bandwidth on the path between a client and server is insufficient to determine if a given request can be satisfied. This is because the available bandwidth changes dynamically and might be either greater or lower in the future (due to unpredictable mobility-induced changes in the network structure and network traffic). In order to contend with the dynamic nature of a C2P2 network, flexible but simple admission control policies must be designed that consider different kinds of information about the current state of the network, and try to predict/estimate based on this information whether the required resources for a given request can be provided. In our admission control framework, we develop two kinds of metrics: a Request Metric (R_m), and an Information Metric (I_m). The request metric R_m (which is normalized to be a number between 0 and 100) is a measure of the amount of resources required for a given request. Hence it depends on the QoS parameter of interest. In particular, we define R_m as the ratio of the requested bandwidth over the nominal maximum wireless link bandwidth B_{max} i.e. $R_m = \frac{\frac{size(X)}{T_X}}{B_{max}} \times 100$ where $size(X)$ is the size of the clip in Mb (MB), and B_{max} is also defined in Mbps (MBps).

While R_m is not policy-dependent, the information metric I_m is, and represents quantitatively the information obtained by the client about the current and future resources available in the network. The information metric is normalized to be a number between 0 and 100. I_m is monotonic in the estimated resource availability — the higher the value of I_m, the more likely it should be that a given request can be satisfied. Note that both I_m and R_m have no units. Now the admission control decision is simply a matter of comparing the two metrics: I_m and R_m, and deciding whether the given I_m predicts sufficient availability of resources for the request with metric R_m. A simple yet flexible form for the admission control policy is simply to take the difference between I_m and R_m and compare with a threshold θ. Thus when $I_m - R_m > \theta$, a policy admits the corresponding request. Otherwise, the request is rejected.

Note that based on our definitions, the threshold can potentially take on values from -100 (which corresponds to allowing all requests) to 100 (which corresponds to rejecting all requests). Thus if θ is too low, there is a danger of greater unsatisfied requests due to bandwidth contention. On the other hand, if it is too high, the number of satisfied requests will be low because too few requests are accepted. The optimal choice of threshold may be an intermediate value and could be scenario-dependent as we shall see in our experiments. Of course, the optimal choice of threshold is also policy-dependent, since I_m is defined differently for each policy. When a request is issued by a client, a policy must either accept or reject requests using the threshold computation.

Content replication and switched proximate server selection are important for robust delivery of high-rate content. The client always chooses the closest server from S candidate servers. When C2P2$_d$'s path to a server breaks at time T_Z, C2P2$_d$ might have downloaded Z bytes of a clip X, $0 \leq Z \leq size(X)$. The bandwidth required to download the remainder of a clip is $\frac{size(X)-Z}{T_X - T_Z}$. If this bandwidth exceeds the wireless network bandwidth, the request is discarded as a failed request. Otherwise, C2P2$_d$ identifies another reachable server and begins downloading the remainder of the clip from that server.

The Mobility-based ADmission control (MAD) policy tries to leverage the mobility pattern in the network by estimating the duration of time a server s_i will be in the radio range of the requesting client C2P2$_d$. This depends on both the direction and speed of s_i and C2P2$_d$. C2PD$_d$ sends out a message (probe) with a certain time to live to obtain the list of proximate servers s_i. Assuming $T(s_i, C2P2_d)$ denotes the duration of time s_i and C2P2$_d$ are expected to be in radio range, for MAD: $I_m = \frac{\sum_{i=1}^{S} T(s_i, C2P2_d)}{\hat{N} T_X} \times 100$. Here, $\hat{N}$ is the total number of nodes reachable from the client via the probe. Note that both the numerator and denominator are in units of time.

This policy only considers the time that a server is within radio range of C2P2$_d$. This is conservative because it ignores the time the servers are reachable via multiple hops.

3. PERFORMANCE EVALUATION

In this section, we quantify the performance of MAD and its comparison to the no-admission control case. We conducted many experiments with different parameter settings. In order to summarize the lessons learnt, we developed utility models to summarize the QoS observed with each policy in one number. These models are presented in Section 3.1. Section 3.2 presents our experimental environment. Section 3.3 demonstrates superiority of MAD to an environment without admission control.

3.1 Utility models for QoS

This section describes three utility models to quantify the QoS metric.

Recall that an admission control policy divides the total number of requests into three groups: R rejected requests, A_f failed admitted requests (due to unsatisfied requirements), and A_g successfully served admitted requests. Different users may value each of these differently, as per their utility model. For the purpose of evaluating our experimental results we consider three distinct utility models that are representative of different levels of QoS that can be provided in a C2P2 ad hoc network. In each model M, we define the joint utility U as a weighted sum:

$$U_M = w_R(M)R + w_{A_g}(M)A_g + w_{A_f}(M)A_f$$

where $w_i(M)$ indicates the weight of component i in model M.

Table 1 shows the weights for three QoS models employed to evaluate the experimental results. The economy model does not penalize a policy for either rejected or failed admitted requests and only rewards accepted requests. The Standard model is indifferent to rejections, but penalizes failed admitted requests as much as it rewards successfully served admitted requests. The Premium model resembles a high QoS environment. It penalizes rejected requests (albeit slightly), rewards admitted requests that are satisfied, and highly penalizes admitted requests that fail. The high penalty for a failed admitted request reflects our intuition that users

Models	Weight of a rejected request, w_R(M)	Weight of a successfully admitted request, w_{A_g}(M)	Weight of a failed admitted request, w_{A_f}(M)
Economy	0	1	0
Standard	0	1	-1
Premium	-0.5	1	-5

Table 1: Three utility models to quantify quality of service with alternative policies.

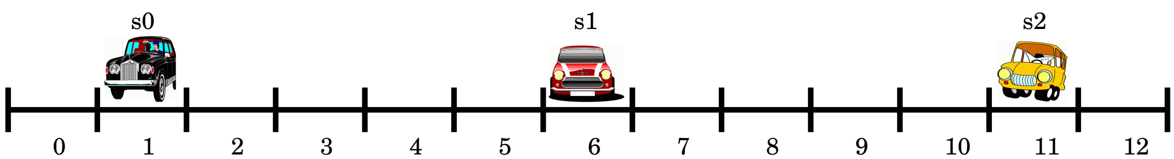


Figure 1: Thirteen road stretches with 3 stationary servers.

are likely to be annoyed if their C2P2 device initiates a download, spends a considerable amount of time (maybe T_X time units) only to discover that it cannot display a clip because an incomplete file is downloaded.

While these utility models can be applied on a per-user basis and different QoS levels can be integrated within the same system in practice, for the purposes of analyzing our experimental result and comparing the different admission control policies, we shall apply each QoS utility model to the system as a whole for the duration of the experiment.

3.2 Experimental Environment

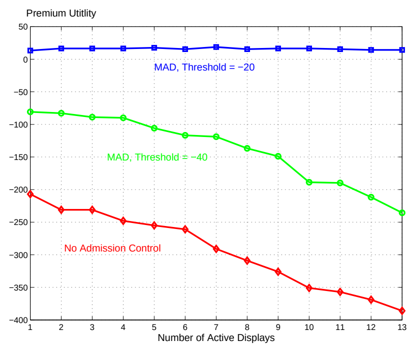
The results presented in this section are based on a new simulator for C2P2 networks written using the C$^{\#}$ programming language. The simulator models road stretches and cars that navigate these stretches. A car might be configured with a C2P2 device that provides a fixed amount of network bandwidth. Each C2P2 device implements the DSR [3] routing policy. However, we believe that a choice of the routing protocol does not impact the trends in the observed results. A proactive protocol like DSDV [7] would have shown similar results. We analyzed different scenarios consisting of a different number of road-stretches, different number of C2P2-equipped cars, mobility patterns, and request specifications.

Our basic experiment consists of thirteen bi-directional road-stretches, numbered from 0 to 12, see Figure 1. Three stationary servers (these could be parked vehicles, access points, or base stations) are located on the following road-stretches: 1, 6, and 11, numbered s_0, s_1, and s_2, respectively. The radio-range of each C2P2 spans 3 road-stretches: its current road stretch and its two adjacent road stretches. Thus, a C2P2 on either road-stretch 0 or 2 is in the radio range of s_0 which is on road-stretch 1. Note that there exists four road stretches that are dark because they are not 1-hop reachable by a server. However, given sufficient number of clients and transitive routing of packets across these clients, these dark road stretches might lit-up as they become network reachable to either one or two servers. We analyzed configurations consisting of 13 possible client C2P2 nodes (in addition to the 3 server nodes), with the simultaneous active displays ranging from 1 to 13 in our experiments. Initially, all cars are assigned to the road-stretches such that they are evenly spaced across the road-stretches. For example, with 13 cars, one car is assigned to each road-stretch. The initial direction of each car is chosen randomly (i.e to move left or right). The speed of each car is fixed at 5 meters per second. Once a car reaches the end of either road stretch 0 or 12, it switches directions and moves toward the opposite end. Within each configuration, the cars that are requesting clips (corresponding to the number of active displays) are chosen randomly and they re-issue requests periodically every 60 seconds.

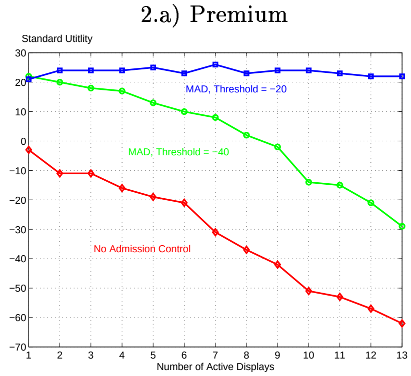
A total of 100 requests are issued in each experiment. This means that with 10 active displays, each client issues 10 requests on average. All three servers are assumed to contain a replica of the clips being requested. Rejected requests do not consume any network bandwidth and are terminated immediately. Once a request is admitted to the system, it downloads a clip for 60 seconds. The maximum network bandwidth of a C2P2 device is assumed to be 10 Mbps. The referenced clip is a media clip with a bandwidth requirement of 340 Kbps. The display time of each clip is fixed at 12 minutes. We require the clip (size ˜30MB) to be downloaded in 60 seconds, requiring a download bandwidth requirement of 4 Mbps.

3.3 A Case for Admission Control

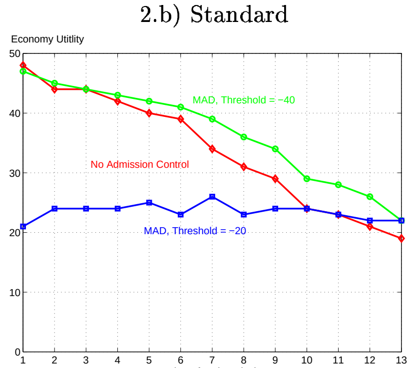
We start with experimental results that justify the use of intelligent admission control policies in an ad-hoc network of C2P2 devices. Figure 2 shows utility of MAD with alternative models when compared with an environment that does not utilize an admission control policy. Figure 3 shows the number of rejected, satisfied, and unsatisfied requests for both MAD and the no admission control case. Without admission control, the



2.a) Premium



2.b) Standard



2.c) Economy

Figure 2: A comparison of Instantaneous-MAD with an environment that employs no admission control. The environment consists of 16 C2P2 devices: 3 stationary servers and 13 mobile clients.

simulator admits a request upon its arrival independent of server availability. Figure 2 shows the performance of MAD for two different thresholds: -20 and -40. The x-axis of this figure shows the number of concurrent clients that initiate the display of a clip, termed number of active displays. This controls the load in the environment and is increased from 1 to 13. The y-axis of this figure is the utility of each model. We present results for all models shown in Table 1. Each presented data point is an average of results obtained from 10 experiments utilizing different seeds. (The random seed impacts the order and identity of clients issuing requests.)

With 3 stationary servers and a maximum wireless bandwidth of 10 Mbps per C2P2 device, our experimental environment supports a total bandwidth of 30 Mbps. All active clients issue their requests at the same time. A total of 100 requests are issued by all clients in this environment. This implies that with two or fewer number of active displays, if there were either no dark regions or unpredictability due to mobility, without admission control, a total of 100 requests would be satisfied.

Figure 3.b shows that only one half of requests are served successfully with either 2 or fewer active displays. This is because of mobility and dark regions that cause an active client to starve for data, failing to download a clip X in T_X time units. This is reflected in the number of unsatisfied requests admitted to the system, see Figure 3.c. A primary observation from Figures 2 and 3 is that an environment with no admission control is clearly inferior to Instantaneous-MAD with θ=-40.

With the Premium model, MAD is superior by several orders of magnitude. With the Standard utility model, MAD remains superior. With the Economy model, however, no admission control outperforms MAD when θ=-20. This threshold renders MAD too conservative, forcing it to admit too few requests.

While all these requests are processed successfully, see Figure 3.c where A_f=0 with θ=-20, MAD does reject requests that can be processes successfully. With a more relaxed threshold (-40), these requests are admitted into the system, enabling MAD to outperform the environment with no-admission control when using the Economy model.

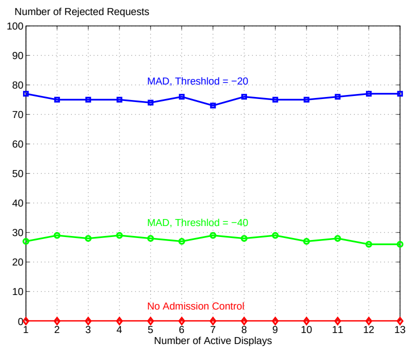
Note that utility of MAD with θ=-20 is a constant positive with the Economy, Standard, and Premium models. When θ=-40, MAD's utility drops as a function of load because the number of unsatisfied requests increases with this threshold, see Figure 3.c.

4. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

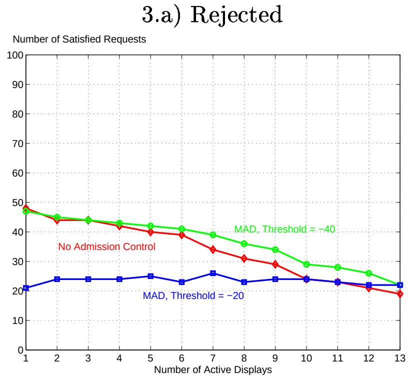
The primary contribution of this paper is a mobility based admission control policy (MAD) to download a continuous media clip in an ad-hoc network of C2P2 devices. Our experimental results demonstrate an environment that employs admission control is superior to one without admission control. MAD can be extended to consider the available bandwidth at each server. We are currently running extensive simulations for such a policy. An obvious research direction is to explore the streaming of a clip across the C2P2 devices. This adds an additional metric which considers the number of hiccups and their duration.

5. ACKNOWLEDGMENTS

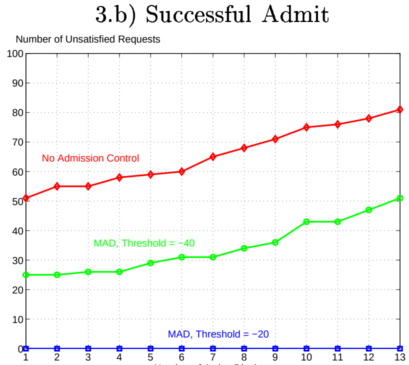
This research was supported in part by NSF grant IIS-0307908, National Library of Medicine LM07061-01, and an unrestricted cash gift from Microsoft Research.



3.a) Rejected



3.b) Successful Admit



3.c) Failed Admit

Figure 3: Rejected (R), successfully admitted (A_g), and failed admitted (A_f) requests with different number of active displays and 16 C2P2 devices: 3 stationary servers, and 13 mobile clients.

6. REFERENCES

[1] S. Ghandeharizadeh and T. Helmi. An Evaluation of Alternative Continuous Media Replication Techniques in Wireless Peer-to-Peer Networks. In *Third International ACM Workshop on Data Engineering for Wireless and Mobile Access (MobiDE, in conjunction with MobiCom'03)*, September 2003.

[2] S. Ghandeharizadeh, B. Krishnamachari, and S. Song. Placement of Continuous Media in Wireless Peer-to-Peer Networks. *IEEE Transactions on Multimedia*, April 2004.

[3] D. B Johnson and D. A Maltz. Dynamic Source Routing in Ad Hoc Wireless Networks. In Imielinski and Korth, editors, *Mobile Computing*, volume 353. Kluwer Academic Publishers, 1996.

[4] E. Knightly and N. Shroff. Admission Control for Statistical QoS: Theory and Practice. In *IEEE Network, vol. 13, no. 2, pp. 20–29*, 1999.

[5] S. Lee, G. Ahn, X. Zhang, and A. Campbell. INSIGNIA: An IP-Based Quality of Service Framework for Mobile ad Hoc Networks. *Journal of Parallel and Distributed Computing*, 60(4):374–406, 2000.

[6] E. Pagani and G. Rossi. A framework for the Admission Control of QoS multicast traffic in mobile ad hoc networks. In *Proceedings of the 4th ACM international workshop on Wireless mobile multimedia*, pages 2–11. ACM Press, 2001.

[7] Charles E. Perkins and Pravin Bhagwat. Highly dynamic destination-sequenced distance-vector routing (dsdv) for mobile computers. In *Proceedings of the conference on Communications architectures, protocols and applications*, pages 234–244. ACM Press, 1994.

[8] M. Rahin and M. Kara. Call Admission Control Algorithms in ATM Networks: A Performance Comparison and Research Directions. In *Interim Research Report Draft*, 1998.

[9] P. Sinha, R. Sivakumar, and V. Bharghavan. CEDAR: a Core-Extraction Distributed Ad Hoc Routing Algorithm. In *INFOCOM (1)*, pages 202–209, 1999.

[10] K. Wu and J. Harms. QoS Support in Mobile Ad Hoc Networks. In *Crossing Boundaries- an interdisciplinary journal, vol 1, no 1*, Aug 1998.

[11] H. Xiao, W. Seah, A. Lo, and K. Chua. A Flexible Quality of Service Model for Mobile Ad-Hoc Networks. In *IEEE VTC*, 2000.

[12] C. Zhu and M. Corson. QoS routing for mobile ad hoc networks. In *IEEE Infocom*, June 2001.