Towards End-to-End QoS in Ad Hoc Networks Connected to Fixed Networks

David Remondo
Telematics Engineering Dep., Catalonia Univ. of Technology (UPC)
Av. del Canal Olímpic s/n
08860 Castelldefels (Barcelona), SPAIN
Tel. +34 93 413 7053 E-mail: remondo@mat.upc.es

Abstract

We evaluate the scalability of a new protocol, named DS-SWAN, designed to support end-to-end QoS in ad hoc networks connected to fixed networks that use DiffServ. When congestion is excessive for the correct functioning of real-time applications, DS-SWAN determines the source of the problem and, if it is the case, allocates more resources to high priority traffic in the ad hoc network. The analysis includes the scalability with respect to the number of real-time traffic sources and node mobility in addition to the impact of best-effort traffic load. Simulation results show an improvement of end-to-end delays and jitter for real-time flows without starvation of background traffic.

1. Introduction

Providing QoS in wireless ad hoc networks is a challenging research area. The network's ability to provide QoS depends on the characteristics of all network components, from transmission links to the MAC and network layers. In these networks, links have a relatively low, highly variable capacity and high loss rates. Besides, node mobility provokes frequent link breakages and link layers typically use unlicensed spectral bands. Therefore, present QoS architectures for wired networks are unsuitable.

A recent overview on QoS support in wireless ad hoc networks can be found in [14]. Important QoS components include QoS aware medium access control, QoS oriented routing and resource-reservation signalling. Among numerous MAC protocols and improvements that have been proposed, a protocol that can provide QoS guarantees to real-time traffic in a distributed wireless environment is Black-Burst (BB) [16]. This protocol is built upon IEEE 802.11 DCF and has good QoS characteristics as far as the traffic flows have constant bit rates. An overview of proposed modifications to IEEE 802.11 for QoS support at MAC level, specifically providing traffic differentiation, can be found in [2].

QoS routing refers to the discovery and maintenance of routes that can satisfy QoS objectives under given resource constraints, while QoS signalling is responsible for flow admission control and resource reservation along the established route. INSIGNIA is the first QoS signalling protocol designed for resource reservation in ad hoc environments [15]. It supports in-band signalling by adding a new option field in the IP header to carry the signalling control information. Like RSVP, the service granularity supported by INSIGNIA is per-flow. If the required resource is unavailable, the flow will be degraded to best-effort service. QoS reports are sent to the source node periodically to report network topology changes, as well as QoS statistics (loss rate, delay, and throughput).

SWAN is an alternative to INSIGNIA with improved scalability properties. SWAN is a stateless network scheme specifically designed for wireless ad hoc networks employing a best-effort distributed wireless MAC [3]. Intermediate nodes do not keep any per-flow information and thus avoid complex signalling and state control mechanisms and make the system more simple and scalable.

There has been little research on the support of QoS when a wireless ad hoc network is attached to a fixed IP network. In this context, co-operation between the ad hoc network and the fixed network can facilitate end-to-end QoS support [1]. In [13], a new protocol, named DS-SWAN (Differentiated Services-Stateless Wireless Ad Hoc Networks), was presented. This protocol comprises co-ordination actions between Differentiated Services (DiffServ) [2] domains in the fixed network and a QoS oriented resource-reservation signalling scheme in the ad hoc network. The signalling scheme is a modified version of SWAN with adaptive parameters. The present work contains performance evaluation results on the scalability of DS-SWAN. We focus on the rather demanding case of improving end-to-end QoS for VBR (Variable Bit Rate) real-time traffic in the presence of background traffic.

The paper is structured as follows: Section 2 describes SWAN, Section 3 presents DS-SWAN, Section 4 shows the performance evaluation results and Section 5 concludes the paper.

2. SWAN

SWAN distinguishes between two traffic classes: real time and best effort. When best-effort packets arrive at a node, they enter a leaky-bucket traffic shaper that has a previously calculated rate, derived from an AIMD (Additive Increase Multiplicative Decrease) rate control algorithm. Every node measures the MAC delays continuously and this information is used as feedback for the rate controller. Every T seconds, each device increases its transmission rate gradually (increment rate of c bit/s) until the packet delays at the MAC layer become excessive. As soon as the rate controller detects excessive delays, it reduces the rate of the shaper with a decrement rate (decrease of r %). Rate control restricts the bandwidth of best-effort traffic so that real-time applications can use the required bandwidth.

For the real-time traffic, SWAN uses sender-based admission control. This mechanism works by sending an end-to-end request/response probe along the constructed route to estimate the bandwidth availability at each node and then determines whether a new real-time session should be admitted or not.

3. DS-SWAN

Section 3.1 describes the functioning of DS_SWAN for upstream traffic and Section 3.2 describes the functioning of the protocol for downstream traffic.

3.1. DS-SWAN for upstream traffic

We consider a scenario where background traffic and real-time VBR traffic is sent from mobile nodes in the ad hoc network to hosts located in the fixed network (see Fig. 1).

For the real-time traffic, the DiffServ service class is the EF (Expedited Forwarding) PHB (Per-Hop Behaviour).

The number of dropped packets at the ingress edge router and the end-to-end delay of the real-time connections are associated with the QoS parameters of the SWAN model in the ad hoc network. We observe that if the rate of the best-effort leaky bucket traffic shaper is lower, then best-effort traffic is more efficiently restricted and real-time traffic is not so much influenced by best-effort traffic, thereby maintaining the required QoS.

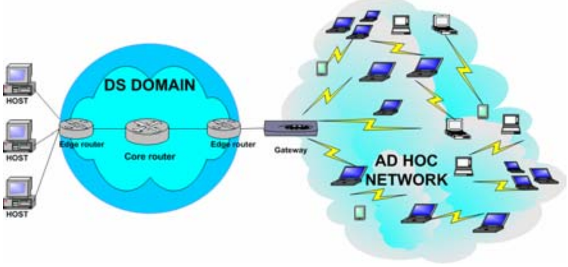

Fig. 1. Considered scenario.

We have selected a relevant real-time application that implies burstiness and that contains end-to-end delay information: VBR Voice-over-IP (VoIP) [4]. The ITU-T recommends in its standard G.114 that the end-to-end delay should be kept below 150 ms to maintain an acceptable conversation quality [5]. Also, for Pulse Code Modulation encoding with the G.711 codec, the packet loss rate should never be larger than 5% [6].

In DS-SWAN, the ingress edge router monitors the number of EF packets that are dropped by its token bucket meter. On the other hand, the corresponding nodes in the fixed IP network periodically monitor the average end-to-end delays of the real-time flows. When a destination node detects that the end-to-end delay of one VoIP flow approached the threshold (i.e. becomes greater than 140 ms), it sends a QoS_LOST warning message to the ingress edge route (see Fig. 2).

We have observed from initial simulation runs that the number of dropped VoIP packets in the ad hoc network is usually well below 1% when DS-SWAN is used. Therefore, we establish that if the number of dropped VoIP packets at the ingress edge router is less than 4% and this router has received a QoS_LOST message, then it sends the QoS_LOST message to the ad-hoc network to inform that the system is too congested to maintain the desired QoS (in this case, it is due to excessive delays at the ad hoc network). On the other hand, when the number of lost packets at this edge router is higher than 4%, it does not make any sense to send QoS_LOST messages to reduce the end-to-end delays.

When the edge router sends a QoS_LOST message to the ad hoc network, it sends the message only to the VoIP sources generating flows that have problems to keep their end-to-end delays under 150 ms, which will obviously also arrive at the intermediate nodes along the routes. All these nodes forward the QoS_LOST message to all their neighbours because they may be contending with them for medium access.

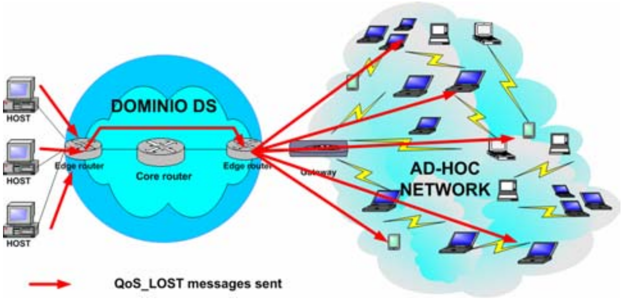

Fig. 2. QoS_LOST messages sent for upstream traffic control.

The nodes in the ad hoc network use priority scheduling at the MAC layer to prioritize routing packets and QoS_LOST packets. When a node in the ad hoc network receives the QoS_LOST message, it will react by modifying the parameter values in the AIMD rate control algorithm. In DS-SWAN, every time that a QoS_LOST message is received, the node decreases the value of c by Δc- bit/s with a certain minimum value. When no QoS_LOST message is received during T seconds, the node increases the value of c by Δc+ bits/s unless the initial value of c has been reached. When a wireless node receives a QoS_LOST message, it also increases the value of r by Δr+ up to a maximum value. When no QoS_LOST message has been received in the period T, the value of r is decreased by Δr- until the initial value of r is reached.

SWAN has a minimum rate m for the best-effort leaky bucket traffic shaper. In DS-SWAN, nodes are allowed to reduce m. When a node receives a QoS_LOST message, it reduces m by Δm- bit/s. However, this parameter value is kept above a minimum value of m_0 bit/s and is increased Δm+ bits/s every T seconds up to the initial value when the mobile nodes do not receive a warning message during T seconds. Therefore, we change the parameter values of SWAN dynamically according to the traffic conditions existing not only in the ad hoc network but also in the fixed network.

Table 1 shows the specific parameter values that we have selected in the simulations presented in Section 5.

It is important to notice that some nodes might receive the QoS_LOST message more than once because they are neighbours of two intermediate nodes so that they will act over the Leaky Bucket parameters several times. This is justified because this node is contending for medium access with several nodes congestion problems.

Table 1. Parameter values (upstream traffic).

Initial value of c	41 Kbit/s
Δc-	10 Kbit/s
Δc+	50 bit/s
Minimum value of c	11 Kbit/s
Initial value of r	50 %
Δr+	10 %
Δr-	1 %
Maximum value of r	90 %
Initial minimum rate	31 Kbit/s
Δm-	10 Kbit/s
Δm+	50 bit/s
m0	11 Kbit/s

3.2. DS-SWAN for downstream traffic

The control of traffic sent in this direction is especially interesting because now the gateway, which is carrying all the flows, is contending for medium access with nodes in the ad hoc network under the same conditions and it could easily become a bottleneck.

The main difference of the functioning of DS-SWAN in this case is the role assigned to certain network elements. The following differences should be emphasized:

- The ingress edge router is now the nearest to the hosts located in the fixed network.
- The destination nodes in the ad hoc network are now responsible for measuring the end-to-end delays of the packets.

If the end-to-end delays measured by the destination nodes are excessive, these nodes will send a QoS_LOST message to the egress edge router (the nearest to the gateway), which will forward it to the ingress edge router. This router will check whether the VoIP traffic losses are lower than 5% (see Fig. 3); if it is the case, the ingress edge router will inform nodes in the ad hoc network by sending them a QoS_LOST message across the DS domain (see Fig. 4).

With the aim that the QoS_LOST messages do not increase the end-to-end delays of VoIP traffic, now the QoS_LOST messages are sent in both directions and more spaced in time. Furthermore, the parameters of the DS-SWAN protocol are also different in this context, as given in Table 2.

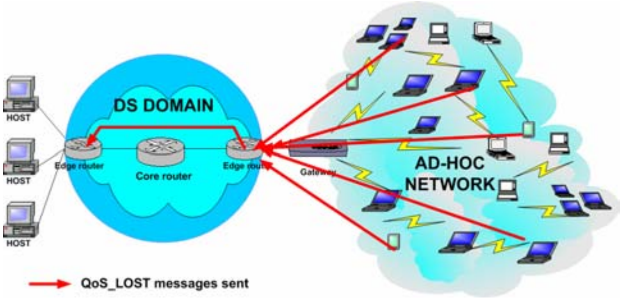

Fig. 3. QoS_LOST messages sent by the destination nodes.

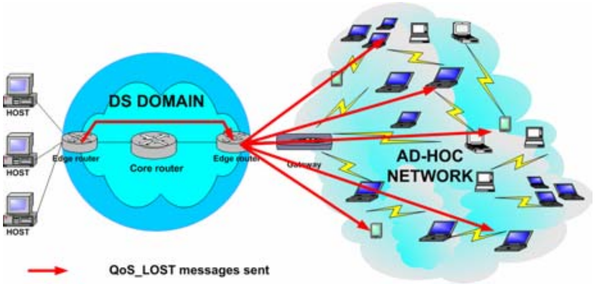

Fig. 4. QoS_LOST messages sent by the ingress edge router.

Table 2. Parameter values (downstream traffic).

Initial value of c	41 Kbit/s
Δc-	15 Kbit/s
Δc+	50 bit/s
Minimum value of c	11 Kbit/s
Initial value of r	50 %
Δr+	20 %
Δr-	1 %
Maximum value of r	90 %
Initial minimum rate	31 Kbit/s
Δm-	20 Kbit/s
Δm+	50 bit/s
m0	11 Kbit/s

4. DS-SWAN performance evaluation

To investigate the performance of DS-SWAN with a relatively realistic physical-layer model, simulations were carried out with the NS-2 tool [9].

We consider a single DiffServ domain (DS-domain) between the corresponding hosts and two wireless gateways. The basic scenario consists of 20 mobile nodes, 2 gateways, 3 fixed routers and 3 corresponding hosts. We consider a hybrid gateway discovery method [10] for gateway selection. The mobile nodes are uniformly distributed in a rectangular region of 700 m by 500 m. Each mobile node selects a random destination within the area and moves toward it at a velocity uniformly distributed between 0 and 3 m/s. Upon reaching the destination, the node pauses for 20 s, selects another destination and repeats the process. The links are IEEE 802.11b.

For upstream traffic, background CBR (Constant Bit Rate) traffic is generated by 13 of the mobile nodes, while VBR VoIP traffic is generated by 15 of the mobile nodes. The destinations of each of the background and VoIP flows are chosen randomly among the three hosts in the wired network.

We consider CBR as background traffic instead of TCP. The reason is that TCP performs poorly in ad-hoc networks [11]. On the contrary, many authors (e.g., [6] and [12]) use CBR as background traffic. However, we have carried out complementary simulations which prove the good performance of DS-SWAN using TCP as best-effort traffic.

The VoIP traffic is modelled as a source with exponentially distributed on and off periods with 1.004 s and 1.587 s average each. Packets have a constant size and are generated at a constant inter-arrival time during the on period. The VoIP connections are activated at a starting time chosen from a uniform distribution in [10 s, 15 s].

Background traffic has a rate of 48 Kbit/s and a packet size of 120 bytes. To avoid synchronization, the CBR flows have starting times chosen randomly from the interval [15 s, 20 s] for the first source, [20 s, 25 s] for the second source and so on.

Fig. 5 shows the end-to-end delay for VoIP traffic with SWAN and DS-SWAN for upstream traffic. Using SWAN, the end-to-end delays increase progressively because the system becomes congested with an increasing amount of background traffic. With SWAN, end-to-end delays are too high for an acceptable conversation quality [5]. In DS-SWAN, the end-to-end delays of the VoIP flows are kept below the required 150 ms.

The aggregate throughput of background traffic in DS-SWAN was reduced with respect to SWAN by 30% at most. The packet loss rate for VoIP was well below the required 5% in all simulations.

In order to analyze the scalability of the DS-SWAN protocol with respect to the number of VoIP sources, we carried out simulations with two different networks:

- A network of small size (20 nodes) that has 13 CBR sources within an area of 700 m x 500 m. The gateways are located at co-ordinates (100 m, 250 m) and (600 m, 250 m). This case is denoted as “DS-SWAN-20 nodes”.
- A network of larger size (40 nodes) that has 26 CBR sources and double area (990 m by 707 m). The gateways are located at (141 m, 354 m) and (849 m, 354 m). This is denoted as “DS-SWAN-40 nodes”.

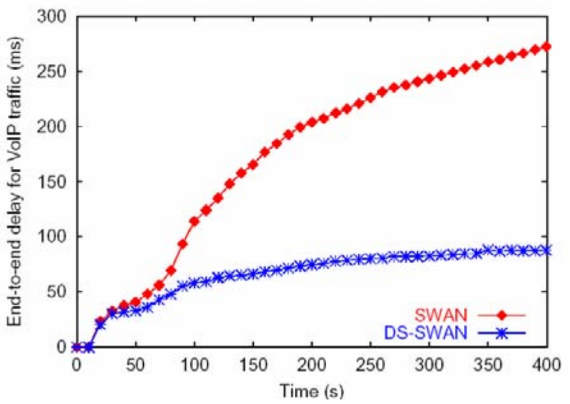

Fig. 5. End-to-end delay for VoIP traffic: SWAN vs. DS-SWAN.

The results, shown in Fig. 6, represent the maximum end-to-end delay for VoIP traffic for both networks with respect to the number of VoIP sources. These parameter values have been measured from the second 80 on, when all best-effort and VoIP sources are active (stationary conditions.

In the smaller network, we can observe that when the number of VoIP sources becomes larger than 4, the DS-SWAN protocol start to act, reducing the end-to-end delays of the VoIP flows. In both cases, the maximum end-to-end delays for VoIP traffic were below than 150 ms and best-effort traffic did not undergo starvation. In all simulation runs the total number of VoIP traffic losses was around 1 % at most.

To analyze the scalability of DS-SWAN with respect to mobility, we have run simulations with different pause times. We consider two different networks:

- A network of small size (20 nodes) that has 13 CBR sources in the ad-hoc network and an area of 700 m x 500m. The gateways are located at (100 m, 250 m) and (600 m, 250 m).
- A network of larger size (40 nodes) that has 26 CBR sources (of best-effort traffic) in the ad-hoc network and an area of 990 m x 707 m. The gateways are located at (141 m, 354 m) and (849 m, 354 m).

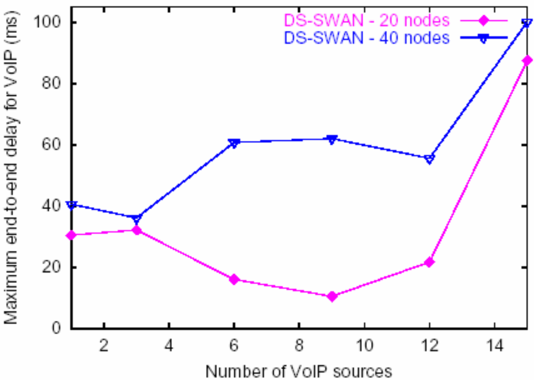

Fig. 6. Maximum end-to-end delay for VoIP traffic as a function of network size and number of real-time sources.

Fig. 7 and Fig. 8 show the maximum end-to-end delay for VoIP traffic and the throughput of best-effort traffic for both networks with respect to mobility (pause time). We observe that the maximum end-to-end delays are always kept very stable and low in the smaller network. In the larger network, the maximum end-to-end delays are not significantly influenced by mobility, and they are kept below 150 ms in all cases. On the other hand, we notice that the throughput of best-effort traffic is not sensitive to mobility. In all simulation runs, the maximum packet loss rate of VoIP traffic was below 1 % [6].

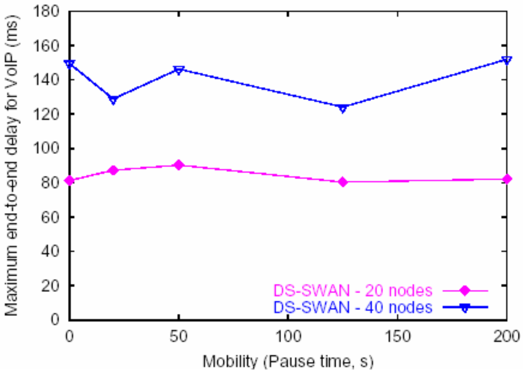

Fig. 7. Maximum end-to-end delay for VoIP traffic as a function of network size and mobility.

Due to space restrictions, we do not show figures on the scalability of DS-SWAN for downstream traffic. The most important results were that the maximum end-to-end delay is below 150 ms for the network sizes and number of VoIP sources employed in the simulation results shown.

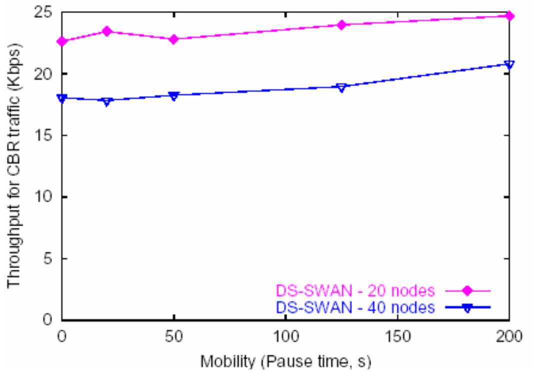

Fig. 8. Throughput of best-effort traffic as a function of network size and mobility.

5. Conclusions

DS-SWAN reduces the average end-to-end delays of VoIP flows and improves the average throughput for best-effort traffic. DS-SWAN is scalable with respect to network size, traffic load and mobility.

If we incorporate the proposed solutions, it will be possible offer better support for real-time application requirements when co-existing with background traffic in the context of ad hoc networks connected to wired IP networks.

Further work is needed for the co-existence of more traffic priorities and traffic with different QoS constraints. Also, the author believes that the co-operation of the presented scheme with suited dynamic routing protocols and already proposed service differentiation schemes at MAC level can yield further performance improvements and, although it will not guarantee QoS levels, it will imply more efficient resource utilization and provide satisfactory service levels with high probability.

6. Acknowledgements

The author would like to acknowledge the extensive and accurate work of Mari Carmen Domingo, thanks to which the simulation results were obtained.

This work was partially supported by the "Ministerio de Ciencia y Tecnología" of Spain under the programme Ramón y Cajal and under the project TIC2003-08129-C02.

References

[1] Y.L. Morgan and T. Kunz, “PYLON: An Architectural Framework for Ad-hoc QoS Interconnectivity with Access Domains,” in *Proc. 36th Annual Hawaii Int. Conf. on System Sciences*, Jan. 2003, pp. 309–318.

[2] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang and W. Weiss, “An architecture for differentiated service,” *Request for Comments (Informational) 2475,* Internet Engineering Task Force, Dec. 1998.

[3] G.-S. Ahn, A.T. Campbell, A. Veres and L.-H. Sun, “SWAN,” draft-ahn-swan-manet-00.txt, *Work in Progress,* Feb. 2003.

[4] D. Chen, S. Garg, M. Kappes and K.S. Trivedi, “Supporting VBR Traffic in IEEE 802.11 WLAN in PCF Mode,” in *Proc. OPNETWORK’02*, Washington D.C., Aug. 2002.

[5] ITU-T Recommendation G.114, “One way transmission time,” May 2000.

[6] P.B. Velloso, M.G. Rubinstein and M.B. Duarte, “Analyzing Voice Transmission Capacity on Ad Hoc Networks,” in *Proc. Int. Conf. on Communications Technology - ICCT*, Beijing, China, Apr. 2003.

[7] R. Wakikawa, J. T. Malinen, C. E. Perkins, A. Nilsson, and A. J. Tuominen, “Global connectivity for IPv6 mobile ad-hoc networks”, Internet Engineering Task Force, *Internet Draft* (Work in Progress), July 2002.

[8] C.E. Perkins, E.M. Royer, “Ad-hoc On-demand Distance Vector routing,” in *Proc. of the 2nd IEEE Workshop on Mobile Computing Systems and Applications,* New Orleans, U.S.A., Feb. 1999.

[9] NS-2: Network Simulator, *http://www.isi.edu/nsnam/ns*.

[10] P. Ratanchandani and R. Kravets, “A hybrid approach to internet connectivity for mobile ad hoc networks”, *Proceedings of WCNC 2003*, Vol. 3, New Orleans, Louisiana, USA, March 2003, pp. 1522-1527.

[11] A. Jain, A. Pruthi, R.C. Thakur, and M.P.S. Bhatia, “TCP analysis over wireless mobile ad hoc networks”, *Personal Wireless Communications, 2002 IEEE International Conference on*, New Delhi, India, 15-17 Dec. 2002, pp. 95 – 99.

[12] D. Chen, S. Garg, M. Kappes and K. S. Trivedi, “Supporting VoIP traffic in IEEE 802.11 WLAN with enhanced medium access control (MAC) for quality of service”, *www.research.avayalabs.com/techreport/ALR-2002-025-paper.pdf.*

[13] M.C. Domingo and D. Remondo, “Quality of Service Support in Wireless Ad Hoc Networks Connected to Fixed DiffServ Domains”, in *Proc. Conference on Personal Wireless Communications (PWC 2004),* Delft, Netherlands, *Lecture Notes in Computer Science,* Berlin, 2004, Springer Verlag.

[14] S. Chakrabarti and A Mishra, “Quality of service challenges for wireless mobile ad hoc networks,” *Wireless Communications and Mobile Computing,* vol. 4, 2004, pp. 129-153.

[15] S. B. Lee and A. Campbell, INSIGNIA, Internet Engineering Task Force, *Internet Draft* (Work in Progress), May 1999.

[16] J.L. Sobrinho and A.S. Krishnakumar, “Quality-of-service in ad hoc carrier sense multiple access wireless

networks," *IEEE Journal on Special Areas in Communications,* vol. 17 (8), 1999.