Experimenting with Fuzzy Logic for QoS Management in Mobile Ad Hoc Networks

Lyes Khoukhi and Soumaya Cherkaoui, Lyes.Khoukhi@USherbrooke.ca

University of Sherbrooke, Department of Electrical and Computer Engineering, QC, Canada

Summary

This paper describes a semi-stateless approach based on a fuzzy logic system for wireless mobile ad hoc networks. The proposed model, called FuzzyMARS, assures service differentiation delivery with low delay transmission. FuzzyMARS integrates a set of mechanisms: a fuzzy logic system for best-effort traffic regulation, three schemes for real-time traffic regulation, and an admission control mechanism for real-time traffic control. The architecture supports both real-time UDP traffic and best-effort UDP and TCP traffic. The use of fuzzy logic in wireless ad hoc networks, as simulations show, can add more flexibility and capability when operating with imprecise or instable information collected in an dynamic topology. The performance evaluation of the proposed model was studied under different mobility, channel, and traffic conditions. The results of simulations confirm that the proposed model can achieve a low and stable end-to-end delay under different network scalability and mobility conditions. Fuzzy logic techniques promise to be efficient tools for reducing delay transmission of multimedia applications in wireless ad hoc networks and deserve further attention and study.

Key words:

Mobile Ad hoc Networks, QoS Support, Fuzzy Logic, Traffic Regulation, Real-time, Admission control.

1. Introduction

Current accelerated developments in wireless and computer technology will soon allow ad hoc networks to become practicable and valuable in a wide variety of applications. Session-based real-time applications such as multimedia voice conversations are among the most desired future applications of ad hoc networks, and the evolution of multimedia technology and its commercial interest to reach the wide public have made Quality of Service (QoS) in wireless mobile ad hoc networks an avoidable task. QoS assures the guaranty by the network to satisfy a set of predetermined service requirements for the user in terms of end-to-end delay, available bandwidth, and so on. QoS support is of central of importance in determining the success of the network-user relation.

Providing multimedia service with QoS guarantee in wireless ad hoc networks is more complex than in fixed-IP networks. The characteristics of an ad hoc network present a significant challenge; this is because of the limited bandwidth resource, dynamic topology, distributed multi-hop communication, and shared wireless medium. Several woks (see Section II) have been dedicated to address these issues. The existing QoS approaches can be classified into “stateful” and “stateless” approaches according to the implemented mechanisms used to support QoS. It is observed that the stateful approaches (e.g., [1], [3]) need for signalling and complex control mechanisms in order to refresh and update per-flow state information. These schemes need flow state information to be maintained in the network nodes, which is both difficult to manage in highly dynamic networks, and poorly scalable when the number of mobile nodes grows. On the other hand, the stateless approaches (e.g., [2]) choose not to use and maintain flow state information. Rather, they typically use feedback-based mechanisms and local control to support service differentiation and real-time services. However, in the case of relatively large networks with high dynamics, the stateless source-based approaches may get an old view of the real status of resources (for instance, bandwidth). This may cause the problem we call “illusory readings”, in which multiple source nodes read the state of the network simultaneously. This problem can lead to the admission of more traffic than what an intermediate node can really support.

In this paper, we detail a fuzzy logic QoS approach for wireless ad hoc network, named FuzzyMARS. The presented approach, first presented succinctly in [4] is a semi-stateless model with service differentiation delivery and fuzzy logic traffic regulation. FuzzyMARS makes uses of fuzzy logic theory for best-effort traffic regulation, and proposes schemes for real-time traffic regulation, and admission control. As shown in the preliminary results depicted in [4], the benefit of the proposed model is the reduction of the average delay of real-time traffic comparatively to other known models. In this paper, we detail a comprehensive study of the model and its mechanisms; specifically the fuzzy logic system for best-effort traffic regulation, three regulation schemes in order to assure the control of real-time traffic, and the admission control mechanism to support real-time traffic and service differentiation delivery. Comprehensive simulation sets show that the model achieves a lower and more stable delay than other counterparts while keeping sensitively the same average throughputs.

FuzzyMARS model does not require the support of a QoS-capable MAC to deliver service differentiation. Rather,

---

Manuscript received August 5, 2008.
Manuscript revised August 20, 2008.

real-time services are built using existing IEEE 802.11 best-effort MAC technology. The admission control with a temporary reservation process (TRP) is used for UDP real-time traffic. The objective of the regulation technique is to adjust dynamically the transmission of traffic according to the network conditions in order to assure a good utilization of resources. The response to the fluctuations is performed by the source nodes, which adjust consequently the transmissions. The proposed FuzzyMARS model is a semi-stateless approach. It tries to take some advantages and overcome some problems of both “stateless” and “stateful” approaches and uses only some minimal information at network nodes.

The rest of this paper is organized as follows. Section II presents a survey of the related works. In Section III, we describe the main mechanisms of FuzzyMARS architecture. Section IV shows the performance results of FuzzyMARS over both SWAN and IEEE 802.11 MAC models under various network simulation conditions. Finally, Section V concludes the paper.

2. Related Works

The QoS satisfaction problem in wireless ad hoc network has been studied by many researchers. A survey of the existed works reveals that there are many sub-problems that are often brought up when attempting to provide QoS in a wireless mobile network environment. The researches in the area of supporting QoS in the MANETs tend to deal with the following issues: the QoS routing issue that address the best and simple way to find a path through the network that is capable of supporting a requested level of QoS. The variation in resources availability issue that deal with the response to the changes in the resources availability, either as a response to the network mobility, or to the variations in the node links characteristics. The QoS maintenance issue that address the rescheduling and the availability of new routes which can support existing QoS commitments when the network topology changes. Other topics such as QoS-based medium access controllers and fairness issues are also very interesting problems.

Recently, there has been research in the area of supporting QoS in MANETs. The works that exist tend to be based on distributed scheduling algorithms that address QoS routing issue, QoS-based medium access controllers, rescheduling when the network topology changes, and fairness issues. The works in [7]-[15] have studied the QoS routing issue. In [7], we have proposed a flexible QoS routing protocol (AQOPC) based on multi-service classes and multi-path scheme. It provides information about the state of bandwidth, end-to-end delay and hop count in the network. It performs accurate admission control and optimal use of network resources by calculating multiple paths and generating the needed service classes to support different QoS user requirements. In [13], a core-extraction distributed ad hoc routing (CEDAR) algorithm is proposed that uses core extraction, link state propagation, and route computation to support QOS in wireless ad hoc networks. In [8], the authors have addressed the problem of supporting real-time communications in a multihop mobile network using QoS routing that permits bandwidth calculation and slot reservation. This protocol can be applied to two important scenarios: multimedia ad hoc wireless networks and multihop extension wireless ATM networks. The ad hoc QoS on-demand routing (AQOR) is discussed in [9], which integrates signaling functions for resource reservation and QoS maintenance at per-flow granularity.

Some works [8], [13] have proposed table-driven routing approaches for QoS support. However, their performances are low compared to reactive approaches, due to the problem of stale route information [14]. A link-state QoS routing protocol for ad hoc networks (QOLSR) was proposed in [10] in the aim of implementing QoS functionality to deal with limited available resources in a dynamic environment. A ticket-based QoS routing protocol was proposed in [11]. This protocol is based on a model which assumes that the bandwidth of a link can be determined independently of its neighboring links. Using the same model, [12] proposes a QoS multi-path routing protocol based on a ticket-distribution scheme to satisfy bandwidth constraints. Unfortunately, this scheme does not consider radio interference problems. The proposed QoS routing protocols can also be classified into two schemes: source routing and distributed routing. Most of the existing distributed algorithms (e.g., [15]) require the maintaining of a global network state at every node, which may cause the scalability problem. On the other hand, the source routing schemes such as [14] suffer from problems of scalability and frequent updates of the state of the network.

It is important to note that the ability to provide QoS depends also on how well the resources are managed at the MAC layer. Some among the works cited above used generic QoS measures and are not tuned to a particular MAC layer [2], [11]. Some others use CDMA to eliminate the interference between different transmissions [8], [16]. The authors in [17] have introduced an on-demand, link-state, multi-path QoS routing protocol which collects information of link bandwidth from source to destination under the CDMA-over-TDMA channel model. Similarly, CDMA-over-TDMA channel model has been adopted in [16] by using the notion of a time slot on a link to calculate the end-to-end path bandwidth. The same model has been used for calculating the end-to-end path bandwidth to develop on-demand QoS routing [8] and DSVD based QoS routing [16].

Other researches presents mechanisms that enable QoS support independent of the routing protocols. The most noteworthy QoS models attempting to establish comprehensive solutions for MANETs are SWAN [2], FQMM [6], and INSINGIA [1]. SWAN proposes a service differentiation in stateless wireless ad hoc networks by using distributed control algorithms. It relies on feedback from the MAC layer as a measure of congestion in the network by using a mechanism of rate control and source-based admission control. It promotes a rate control system that can be used at each node to treat traffic either as real-time or best-effort traffic. However, one of the drawbacks of SWAN is how to calculate the threshold rate limiting any excessive delay that might be experienced. It also uses merely two levels of services: real-time and best-effort traffic.

SWAN and INSIGNIA are intranet QoS models providing services that have to be mapped to either per-flow or per-class services, but SWAN remains the best example of stateless distributed QoS framework developed for wireless ad hoc networks. INSIGNIA is one of the noteworthy QoS frameworks with per-flow granularity and reasonable treatment for mobility. The main goal of INSIGNIA is to provide adaptive QoS guarantees for real-time traffic. It employs an in-band signaling system that supports fast reservation, restoration, and adaptation algorithms. Three levels of services are implemented: best-effort, minimum, and maximum. The bandwidth is the only QoS parameter used in INSIGNIA. FQMM is a hybrid approach combining the advantages of per-class granularity of DiffServ with the per-flow granularity of IntServ. It tries to preserve the per-flow granularity for a small portion of traffic in MANETs, given that a large amount of the traffic belongs to per aggregate of flows, that is, per-class granularity. FQMM offers a good solution for small- and medium-size ad hoc network, but it is not suitable for large networks. We have proposed an intelligent QoS model with service differentiation based on neural networks in mobile ad hoc networks named GQOS [5]. The main objective was to satisfy some QoS requirements, especially the reduction of end-to-end delay, in networks whose topologies change at low to medium rate. GQOS is composed of a kernel plan which assures basic functions of routing and QoS support control, and an intelligent learning plan which assures the training of GQOS kernel operations by using multilayered feedforward neural network (MFNN). The advantages of using neural networks algorithm is the fast learning of different operations performed by the kernel and the reduction of the time processing in the network using the training process.

3. FuzzyMARS Architecture

3.1. Overview of FuzzyMARS Architecture

Figure 1 illustrates the FuzzyMARS architecture. The presented schematic diagram aims to support QoS and to adapt to the dynamic changes of the environment. This is achieved by the cooperation between a set of functionalities and mechanisms integrating a fuzzy logic system. The routing scheme and the temporary resource reservation process perform the discovery of routes and bandwidth reservation as will be discussed in the next section. The admission controller efficiently estimates the local available bandwidth at each node. Many multimedia applications such as VOIP (Voice over IP) and MOIP (Multimedia over IP) are delay sensitive or bandwidth sensitive applications. Hence, providing information about real network state can be useful for the decision of acceptance of new flow. The decision to admit a new flow is done by the admission control mechanism. The classifier is able to differentiate between flows in terms of QoS requirements; best-effort flows and real-time flows, in order to delay the best-effort packets. Note that even the admission control is performed to guaranty enough available bandwidth before accepting new flow; the congestion may be occurred in the network because of mobility of nodes. Therefore, it is of central of importance to assure the traffic regulation. The classified best-effort packets are regulated using a fuzzy logic system according to the application requirements and the network state. The fuzzy best-effort regulation uses the feedback delay received from the MAC layer. The fuzzy regulation process is performed in three steps: fuzzification, inference rules evaluation, and deffuzification. On the other hand, three schemes are proposed in order to assure the regulation of real-time traffic when congestion is observed in the network. In FuzzyMARS model, like in SWAN, it is not necessary to support the QoS-capable MAC to deliver service differentiation. Rather, real-time services are built using existing best-effort IEEE 802.11 MAC technology. In the rest of the paper, we give more details about the proposed architecture mechanisms.

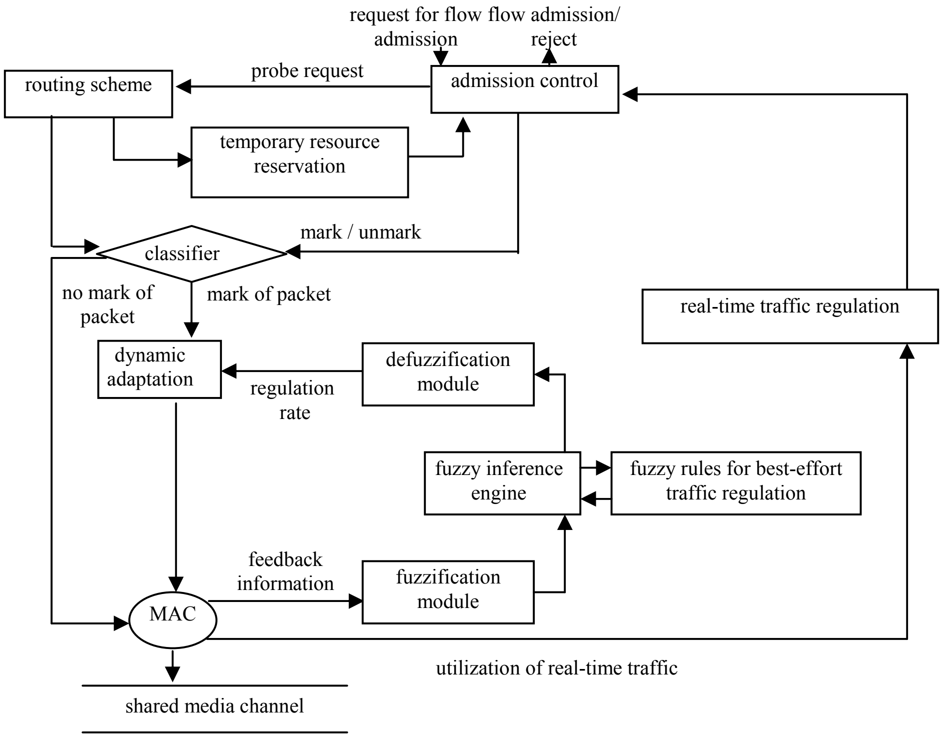


Figure 1: A schematic diagram of FuzzyMARS model

3.2 The admission control mechanism

In this section, we present the admission control mechanism and the process of temporary resource reservation. The main task of the admission control is to determine whether the available resources in a network can meet the requirements of a new flow while maintaining bandwidth levels for existing flows. Accordingly, the decision is performed on the acceptance or reject of flow requirement. This function is conducted together by the source node and other intermediate nodes. Note that in FuzzyMARS, the source node has a final decision to accept or reject the user QoS requirements based on the feedback information about the state of the network. This feedback measure is the packet delay measured by the MAC layer, which is calculated as in [2] by the difference between the time of receiving an ACK packet (from the next-hop) and the time of sending a packet to the MAC layer (from the upper layer). Such in SWAN, the use of a shared wireless channel allows the mobile nodes to listen to packets sent within radio transmission range. This feature allows the admission control to measure the local available bandwidth at each node in the network. The measured available bandwidth is then used by the admission controller to decide if the flow can be admitted for a particular service. The real-time traffic measured by the admission controller is on terms of bits per second. Note that the cooperation between the admission control mechanism and the proposed regulation techniques (Section 3.3 and 3.4) assures that the total traffic (i.e., best-effort and real-time traffic) is bellow a certain threshold rate that would cause congestion and trigger an excessive delay.

In order to avoid the problem of illusory readings, the admission control is performed not only at the source node, but also at the intermediates nodes. The problem of illusory readings is aggravated in the stateless approaches where intermediate nodes do not maintain state information, and admission control is conducted only at the source node in a decentralized manner. This is because multiple source nodes can read simultaneously the state of the network (via request/response probes) and may admit more traffic than what an intermediate node can support. The illusory readings may appear also during the period of network exploration, (i.e. the time between a probe request is sent and a probe response is received at the source node). The source nodes may receive responses to their requests indicating that resources are available when in fact they are not. The following example illustrates the problem of illusory readings. Consider two voice flows at a rate of 32 Kbps and one video flow at a rate of 200 Kbps. If we consider that the available bandwidth at a common intermediate node is 220 Kbps, then only one single video flow could be supported. However, as there is no reservation mechanism at the intermediate nodes, all flows will be initiated. This results in an aggregation of 264 Kbps at the common node. To overcome this problem of local congestion which may cause excessive delay for multimedia application, some stateless approaches such as SWAN use AIMD rate control. However, as mentioned earlier, such approaches do not prevent the illusory readings. The proposed model uses a fuzzy regulation scheme besides the admission controller. The subsections below explain the temporary reservation process (TRP) integrated with the admission controller.

The admission control mechanism first evaluates the available bandwidth in the network, so that the bandwidth requirements of all the flows do not exceed the resources in the network. The estimation of the end-to-end available bandwidth is performed by sending a request from source node toward the destination. For that purpose, an UDP control packet is exploited by using an additional field "*B*" that contains initially the value of the requested bandwidth "B_{req}". At each intermediate node, a comparison is performed between the value of *B* and the available bandwidth "B_{avai}" of the current node. The value of the field *B* is updated if it is bigger than the value B_{avai} of the current node. When the destination receives the UDP control packet, then *B* represents the minimum bandwidth available along the path, and it is copied from UDP to a newly generated short replay message (SRM). The latter packet is transmitted back to the source node, in the same time as the performance of the temporary resource reservation process (TRP). Additional fields are used during TRP mechanism, which are stored in each intermediary node in order to specify the "*temporary reservation status*" of node, the "*status duration*" and the "*flow_identifier*". The first field is set to value of the reserved bandwidth and the status duration is set to a certain value "T". T indicates the period of time within which the temporary reservation is performed. Note that even when the temporary reservation is performed by a flow, other flows can also exploit the available resources of the node. The reserved bandwidth is released just after the expiration of T duration.

In what follows, we evaluate the right status duration to be set at a particular node. The computation of the right status duration needs to take into account the number of hops between the source and the particular node, and also the delays between the intermediate nodes. Let consider Δt the temporary reservation interval of a flow in a given intermediate node. Other flows originating from other source nodes can also use the available resources during Δt. Let μ be the target delivery rate which defines the desired percentage of packets to be sent within the QoS constraint, where $\mu = 1$ corresponds to best QoS guaranty and $\mu = 0$ corresponds to the best-effort transmission. Then, (1) verifies the probability that Δt is bigger than a given time value δ and the flow request to be accepted.

$$P[\Delta t \ \geq \delta] \geq \mu \qquad (1)$$

A good evaluation of (1) requires the destination to be acquainted with the statistical descriptions of delay of each node along the path. However, in many cases, the statistical distribution of such parameter can be approximated by a Gaussian distribution. Under this hypothesis, and assuming independency among nodes statistics, the temporary reservation time among the nodes turns out to be a Gaussian variable. If we consider m_{Tr} and σ_{Tr}^2 the statistical average and variance of the random variable T_r, respectively (T_r is the temporary reservation time in a given node), then the temporary reservation interval statistics can be expressed as follows:

$$P[\Delta t \geq \delta] = 1 - Q\left(\frac{\delta - m_{Tr}}{\sigma_{Tr}}\right) \qquad (2)$$

Where Q represents the complementary distribution function of a Gaussian variable with mean 0 and variance 1.

Let υ be the actual time satisfaction provided by the intermediate node as given by (2). Hence, the flow request would be satisfied even if the average temporary reservation time was decreased to the value m_{Tr} given by (3):

$$m_{Tr} = \upsilon - \sigma_{Tr} Q^{-1}[1 - \mu] \qquad (3)$$

The satisfaction of the requested target delivery rate for a given flow is met if the temporary reservation time is greater than m_{Tr} (m_{Tr} is the time bound of the temporary reservation interval). After the duration m_{Tr} , the temporary reservation status of a node is set to 0. This assures a good utilization of resources, because it permits to the reserved resources to be released in order to be used by other flows.

3.3 Fuzzy logic for best-effort traffic regulation

Fuzzy logic theory [20]-[23] was first introduced as a tool for modeling the uncertainly of natural language, and has been commonly employed for supporting intelligent systems. The Fuzzy logic has proven efficiency in various areas and several applications such as decision support and intelligent control, especially where a system is difficult to be characterized. A fuzzy logic system considers basically three steps: fuzzification, rules evaluation, and deffuzification. The first step is responsible for mapping discrete (called also crisp) input data into proper values in the fuzzy logic space. For that end, membership functions (fuzzy sets) are used to provide smooth transitions from false to true (0 to 1). The second step performs reasoning on the input data by following predefined fuzzy rules. Once the input data are processed by fuzzy reasoning, the deffuzification takes the task of converting back these input data into crisp values.

We propose to regulate the best-effort traffic by using a fuzzy logic system. The use of fuzzy logic can add more flexibility and capability of operating with imprecise information due to the nodes mobility in the wireless ad hoc network. The feedback delays from MAC layer is the key parameter of the proposed strategy which ensures that best-effort traffic coexist well with real-time traffic.

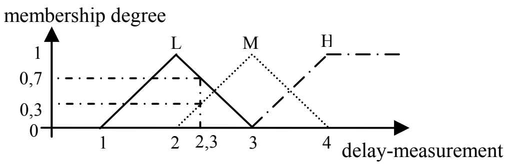
The feedback measure represents the packet delay measured by MAC layer. IEEE 802.11 MAC is used a part of the proposed model. Hence, the measure of the packet delays is performed as follows: at the reception of packet by MAC layer, the later listens to the channel and differs the access to the channel according to the CSMA/CA algorithm. When the MAC gets access to the channel, then RTS-CTS-DATA-ACK packets are exchanged. The reception of ACK packet by the transmitter means that the packet was successfully received by the receiver. The time taken to send the packet between transmitter and receiver including total differed time represent the packet delay. This delay is calculated as the difference between the time that a packet is passed to the MAC layer (from the upper layer), and the time of reception of ACK packet from the receiver. Note that the received packet delay can reflect the network state; a high delay signifies that a possible situation of congestion is occurred in the network. Thus, when one or more packets have greater delays than a certain value, the rate control is triggered in order to reduce the traffic because may be there is a situation of congestion in the network. In FuzzyMARS model, the fuzzy best-effort regulation is combined with the technique of additive increase used in the classic AIMD algorithm. The simulation of the proposed model shows that the fuzzy logic modeling promises to offer an efficient tool for the traffic delay minimization in the wireless ad hoc networks.

FuzzyMARS proposes to control the regulation rate by means of a fuzzy logic system in response to the feedback delay. The regulation process of the best-effort traffic using fuzzy logic is performed in three steps: in the first step, the delay-measurements are transformed into fuzzy sets; then in the second step, a set of fuzzy rules are applied into the fuzzy input in order to compute the fuzzy outputs. The third step translates the fuzzy outputs into crisp values.

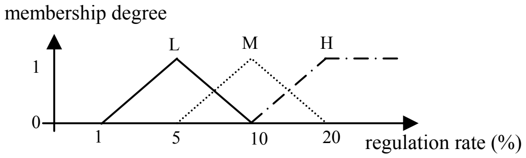
1) Fuzzification: the fuzzy input in our system is the delay-measurement obtained as feedback from MAC layer. The traffic regulation rate represents the fuzzy output. These two parameters have to be converted into fuzzy sets. The fuzzy set is different from an ordinary set, because a fuzzy set may contain elements that have different degree of membership in a set, whereas the elements of an ordinary set are considered members of a particular set if they have full membership in this set [22],[23].

For instance, when the delay-measurement is considered in an ordinary set, then it can only be either low or high and not both simultaneously. However, the delay-measurement in a fuzzy set can be classified as: not high, medium, or quite low. Thus, the membership of an element may be not the same over various fuzzy set. The membership function in a fuzzy set, represented as line or a curve, indicates how to map each input or output parameter onto a membership value. The later is obtained by the process of mapping the parameter value onto a membership function.

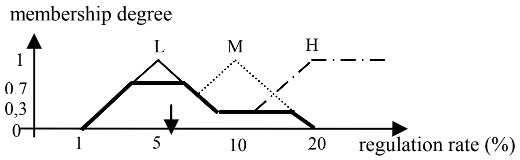
Figure 2 show the different levels that we considered in a fuzzy set: low (L), medium (M), and high (H). The threshold of each level is also defined. Let consider in Figure 2 that the threshold for low delay-measurement is 2 msec, for medium delay-measurement it is 3 msec, and for high delay-measurement it is 4 msec. Then, by mapping the current delay-measurement onto the graph of the membership function, the delay will be allocated with a membership value in each set between 0 and 1. For example, in Figure 2 if the current delay-measurement is 2.3 msec, then this value can be fuzzified into low delay with the degree of 0.7, medium delay with the degree of 0.3 and high delay with the degree of 0. The fuzzification process of the delay-measurement and traffic regulation rate is illustrated in Figure 2.a and 2.b.



2.a) Fuzzy input value of the delay-measurement



2.b) Fuzzy output value of the traffic regulation rate



2.c) Defuzzification process

Figure 2. Fuzzification and defuzzification process

2) Rule evaluation: during this process, a set of fuzzy rules is applied over fuzzy sets by an inference engine. In general, the fuzzy rules are presented as a set of rules "if (…) then (…)". In our study, we use the rule that explains the traffic regulation mechanism:
<< If the delay-measurement is increased, then reduce the actual traffic rate >>.

An increase in the delay-measurement signifies that data packets take more time to be received by the destination node, which means that a possible congestion is occurred. Consequently, the decrease in the traffic transmission has to be performed in the aim of reducing the congestion level occurred in the network. Thus, the decision-making logic of the traffic regulation rate follows the delay-measurement parameter. Note that when the delay-measurement falls into more than one set, more than one decision set can be resulted.

3) Defuzzification: in this step, the fuzzy decision sets obtained are converted into precise quantities. There are several heuristics methods that permit to perform the defuzzification: the mean of maximum, the center of area, and the max criterion [21]. In this study, we have used the mean of maxima (MoM) method [22] as defuzzifier because of its light computational complexity. The evaluation result is obtained as the average of the elements that reach the maximum grade in a fuzzy set.
For example, let consider that the delay-measurement received by a node is 2.3 msec. Then, the regulation rate is determined by using a fuzzy inference process, which is performed in three steps:

1) Fuzzification: Figure 2.a and 2.b show the fuzzification step. The delay-measurement 2.3 msec is fuzzified into medium delay with the degree of 0.3 and low delay with the degree of 0.7.

2) Rule evaluation: a set of rules "if (…) then (…)" is applied in order to determine the fuzzy output.

- if the delay-measurement is low, then the decreasing rate regulation is low,
- if the delay-measurement is medium, then the decreasing rate regulation is medium.

The obtained value (0.7 for low, and 0.3 for medium delay) cut respectively, the set "low" and "medium" of the output parameter.

3) Defuzzification: this step chooses a representative value to the resulted fuzzy set as the final output. For that aim, the mean of maxima is applied. Then, the node performs the regulation of the best-effort traffic according to the representative solution.

3.4 Dynamic regulation schemes

In a wireless mobile ad hoc network, the admission control alone can not assure QoS guarantee since the topology may change after the admission of flows. The network may experience congestion under mobility. Thus, in such situation there is a need to regulate traffic. The objective of the regulation technique is to adjust dynamically the transmission of the traffic according to network conditions in order to assure a good utilization of resources. The source nodes respond to the fluctuations and initiate the adjustments consequently. The detection of overload or congestion at a particular node is possible by periodic measurements of the traffic rate. In the proposed approach, when a congestion is detected, the congested node sends a congestion notification message (CNM) to the source node. Note that this mechanism is different from the one used in other approaches such as SWAN. In the later, it is the destination node who monitors for congestion and notifies the source node using an additional message in a network which is already overloaded. To show the theoretical gain obtained by FuzzyMARS comparatively to SWAN, let consider the scheme in Figure

3; where n is the number of hops between source S and destination D, k is the number of hops between P (the congested node) and S, and t the average transmission time between two hops.

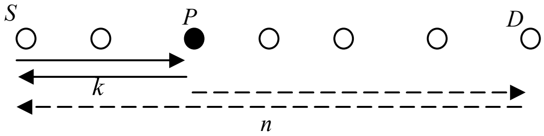


Figure 3: Messages between source, congested node, and destination

The time required by SWAN to notify S is:
$TC_{swan}=(2n-k)\,t$
The time required by FuzzyMARS to notify S is: $TC_{FuzzyMARS} = k\,t$
Hence, the gain of time is shown in (4):
$TCG = TC_{swan} - TC_{FuzzyMARS} = (2n-k)\,t - k\,t$
$TCG = 2\,(n-k)\,t$ (4)

The gain of time obtained is useful in various multimedia applications such as VOIP or video, where the time constraint represents a challenge. The time optimized in (4) permits in one side, to regulate the traffic sent sooner after the detection of congestion, and on the other side to minimize the end-to-end delay. Upon the reception of CNM packet by the source node, the later tries to re-establish the session taking into account both the minimum requirement of the original bandwidth requested and the new state of the network. It is important to note that the sources nodes do not immediately initiate reestablishment upon receipt of a regulate message. Rather, each source node waits for a random amount of time before initiating the reestablishment of traffic. This feature avoid that a number of source nodes simultaneously initiate regulation.

The proposed model assures that the best-effort flows are first reducing their transmission rate (using the fuzzy regulation scheme described in Section III.C) in order to give bandwidth to real-time traffic. The following schemes are proposed to assure the real-time traffic regulation:

1. Flows-based Regulation

In the flow-based regulation scheme (FuzzyMARS-1), the intermediate node detecting the congestion sends a CNM packet to all source nodes which originated the flows. For each source node, the CNM packet includes the traffic regulation rate. Some stateless models such as SWAN, send a notification without specify the rate of flow regulation. The new rates in FuzzyMARS-1 are calculated as follows:

$$\begin{aligned} Cf_1^N &= Cf_1 - h_1 \\ Cf_2^N &= Cf_2 - h_2 \\ &\;.\qquad . \\ &\;.\qquad . \\ Cf_r^N &= Cf_r - h_r \end{aligned} \tag{5}$$

Cf_1, Cf_2… Cf_n are the old rates of the congested sessions, and Cf^N_1, Cf^N_2… Cf^N_n are the new ones. Note that $Cf_1 + Cf_2 + ... + Cf_r > Th$, where r represents the number of flows, and Th represents the threshold admission rate. The distribution of the bandwidth rate over the congested sessions is performed equitably by asking each source j to reduce its traffic rate by a value h_j calculated in (6).

$$h_j = Cf_j\left(1 - \frac{Th}{\sum_{i=1}^{r} Cf_i}\right) \tag{6}$$

2. Priority-Based Regulation

In this scheme (FuzzyMARS-2), rather than considering all congested sessions to have the same priority, we consider different priorities for the flows. In this case, we take into consideration the priority of the flow within the flow sets. The notion of priority is important in many applications and can be useful in several areas. This kind of regulation scheme has not been considered (to the best of our knowledge) in other existing stateless ad hoc models. Note that packets belonging to lower priority flows are selectively dropped prior to packets of higher priority flows. The computations below describe how to compute the new rates of the congested sessions in FuzzyMARS-2:

$$\begin{aligned} Cf_1^N &= Cf_1\left(\frac{\gamma_1 Th}{\sum_{i=1}^{r} \gamma_i Cf_i}\right) \\ Cf_2^N &= Cf_2\left(\frac{\gamma_2 Th}{\sum_{i=1}^{r} \gamma_i Cf_i}\right) \\ &\;.\qquad . \\ &\;.\qquad . \\ Cf_r^N &= Cf_r\left(\frac{\gamma_r Th}{\sum_{i=1}^{r} \gamma_i Cf_i}\right) \end{aligned} \tag{7}$$

$$y_j = \frac{p_j}{\sum_{i=1}^{r} p_i} \tag{8}$$

Where p_1, p_2, …., p_r are the priorities of the different flows, and γ_1, γ_2, …., γ_r the corresponding priority factors calculated by (8). Note that a priority factor γ_j of a flow j is always comprised between 0 and 1.

3. Constraint-based regulation

This scheme of regulation (FuzzyMARS-3) is a way of scheduling between a set of congested flows. This scheduling scheme is realized according to the flow constraints. For instance, if the priority is given to the delay constraint, then the flow with bandwidth constraint can be delayed in the case of congestion. The congested node selects a set of congested flows (Γ) for a period of time (T) and then calculates a new congested set as described in (9).

$$\Gamma(T) = \alpha f_D + \beta f_B + \delta f_J \qquad (9)$$

Where α, β, δ are the ponderation factor which selects the flow to be regulated according to the application constraint and the network state, and f_D, f_B, f_J are the flows traversing the congested node characterized respectively by the constraints of delay, bandwidth, and jitter.

Assuming that Cfg is a set of k flows to be regulated and $\overline{Cfg}$ is the set of the prioritized flows. The flows of the later set are not delayed, which means that they pass, without regulation, through the congested node because of their prioritized constraint. The regulation rate factor h is calculated as follows:

$$h_j = Cf_j \left(1 - \frac{Th}{\sum_{i=1}^{k} Cf_i}\right) \qquad (10)$$

This regulation scheme is useful when a constraint-based preference is considered between the congested flows in the network.

4. Performance Evaluation

In this section, we evaluate the performances of the proposed QoS architecture. The simulator used, is the scalable and efficient NS-2 simulator. Each mobile host has a transmission range of 250 meters and shares an 11 Mbps radio channel with its neighboring nodes. The simulation is realized in two steps: the first one investigates the performance of the proposed model in an environment characterized by a single shared channel. This simulation is the same as the ad hoc model defined in IEEE 802.11 standard. The second simulation considers a multihop environment with different mobility scenarios. The proposed mechanisms implemented in the simulation are: the admission controller, the first regulation scheme of real-time traffic (FuzzyMARS-1), and the fuzzy logic scheme for best effort-traffic regulation.

We compare the network performance of the proposed model with both the 'original model' and the SWAN model described in [2] and. We use the word 'original model' to refer to IEEE 802.11 wireless networks without FuzzyMARS mechanisms. We ran a large set of simulations in order to better understand the characteristics of the proposed model and to study several cases as will be described in what follows.

4.1 Performance of a single shared channel

In this experiment, we consider a single hop simulation environment that consists of a square shape of 150m x 150m. All wireless ad hoc mobile nodes share a single radio channel of 11 Mbps. The source and destination nodes associated with flows are distributed among the mobile nodes in the wireless ad hoc network. The simulation considers a variety of flows types; real-time flows, FTP macro-flows, and WEB micro-flows. The video and voice flows representing real-time traffic are active and monitored for the duration of 100 seconds. Video traffic is modeled as 200 Kbps constant rate traffic with a packet size of 512 bytes. Voice traffic is modeled as 32 Kbps constant rate traffic with a packet size of 80 bytes.

The implemented single hop network considers a multiple scenarios of TCP best-effort traffic, 4 voice and 4 video flows in the aim of better understand the properties of the FuzzyMARS regulation. The TCP traffic is modeled as a mixture of FTP and Web traffic. Web traffic represents micro-flows, whereas FTP traffic corresponds to macro-flows. TCP flows are greedy FTP type of traffic with packet size of 512 bytes. Web traffic is modeled as short TCP file transfers with random file size and random silent period between transfers. The file size is driven from a Pareto distribution with a mean file size of 10 Kbytes and a shape parameter of 1.2. The length of the silent period between two transfers is also Pareto in distribution with the same shape parameter with a mean of 10 seconds.

Figure 4 and 5 show the scalability impact of number of UDP video flows on the average end-to-end delay. Figure 4 show the comparison between FuzzyMARS and the original model, and Figure 5 observes the comparison between FuzzyMARS and SWAN model. The simulation uses a mixture of real-time traffic and TCP best-effort traffic which consists of 16 Web and FTP flows. It is observed in Figure 4 that the original model shows an average end-to-end delay larger than 12 msec with only 5 video flows and over 20 msec with 15 or more video flows.

FuzzyMARS shows delays inferior to 2 msec with 5 video flows and less than 2.5 msec with 20 video flows. Hence, the reduction achieved by FuzzyMARS in terms of the average end-to-end delay is then superior to 85% in comparison to the original model. On the other hand, Figure 5 illustrates that for up to 20 video flows, FuzzyMARS outperforms SWAN by about 8-17%. The average delay in both models grows slowly as the video flows number increases. When the number of video flows becomes high, the average end-to-end delay of traffic in SWAN becomes slightly smaller than in FuzzyMARS. The impact of high number of video flows on the delay is due essentially to the congestion in the nodes. When the packet delay is increased beyond a certain threshold, some flows are dropped in order to maintain low delay for the remaining flows.

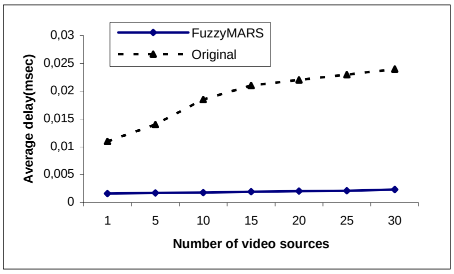


Figure 4: Average delay in the original and FuzzyMARS models vs. number of video flows

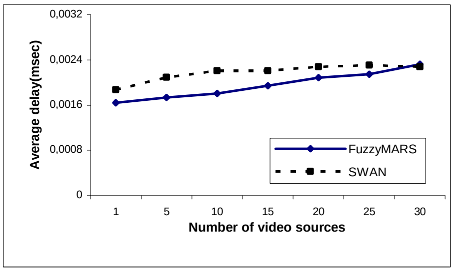


Figure 5: Average delay in FuzzyMARS and SWAN models vs. number of video flows

The impact of the scalability of a growing number of web micro-flows on the average end-to-end delay in FuzzyMARS compared to the original and SWAN models is shown respectively, in Figure 6 and 7. It is observed in Figure 6 that the increasing number of web micro-flows has much more impact on the average delay in the original model than in FuzzyMARS. The average end-to-end delay in FuzzyMARS remains around 2 msec, whereas in the original model the average end-to-end delay grows from 1.8 to 7 msec when the number of web micro-flows increases from 8 to 72 web micro-flows. On the hand, it is observed in Figure 7 that the average delay of traffic in SWAN and FuzzyMARS models is similar for up to 32 web micro-flows. When the number of web micro-flows is smaller than 16, the average delay in SWAN is smaller than that in FuzzyMARS by about 13%. However, for the highest number web micro-flows, the average end-to-end delay of traffic in FuzzyMARS becomes smaller than in SWAN. The gain achieved by the proposed model, comparatively to SWAN model in terms of average end-to-end delay, is about 14-25%.

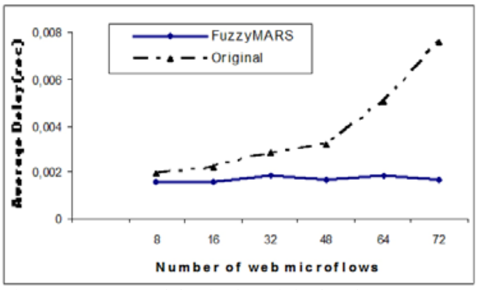


Figure 6: Average delay in the original and FuzzyMARS models vs. number of web micro-flows

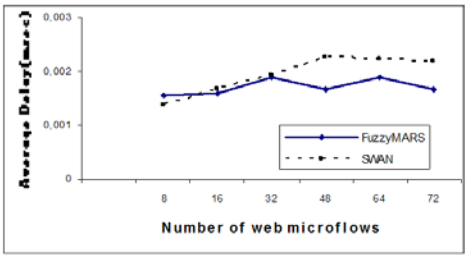


Figure 7: Average delay in FuzzyMARS and SWAN models vs. number of web micro-flows

In the following, we show some illustrations of the trace graph obtained in the simulation. Figures 8-10 show the impact of TCP flows (e.g., FTP macro-flows with web micro-flows) scalability on the average delay in FuzzyMARS in the single hop network. The simulation uses a mixture of 8 real-time flows and TCP flows, which are modeled as in the previous simulations. These Figures show a trace of the packet delays versus the simulation time using respectively, 10, 20, and 30 TCP flows. It is observed that the average end-to-end delay in the proposed model increases as the number of FTP macro-flows and web micro-flows becomes high. For different scenarios of flows scalability in the proposed model, the average delay grows slowly, and the traffic delay becomes almost constant with small and medium number of TCP flows, and it grows only for the highest scalability scenarios. The average delay in Figures 8, 9 and 10 is respectively, 1.96 msec, 2.25 msec, and 2.63 msec. The previous results show that the proposed model provides low delay even in a wireless network with a high number of TCP flows scenarios.

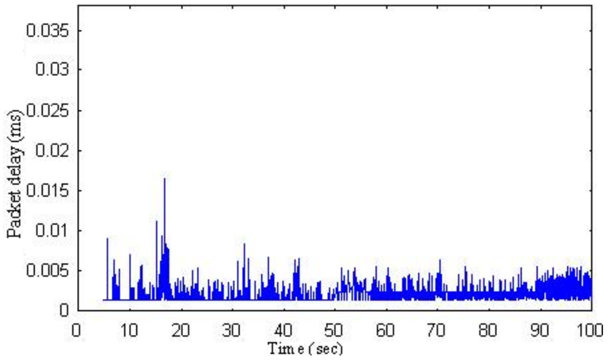


Figure 8: The packets delay in FuzzyMARS with 10 TCP flows vs. simulation time

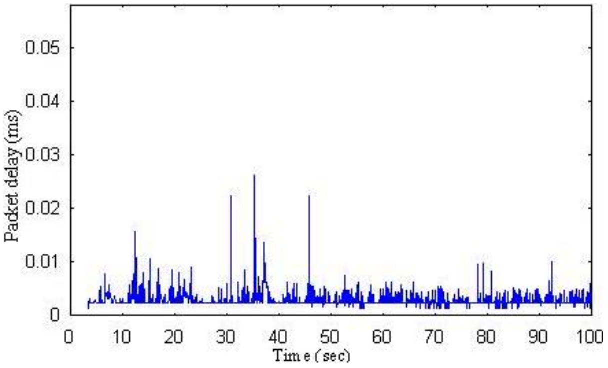


Figure 9: The packets delay in FuzzyMARS with 20 TCP flows vs. simulation time

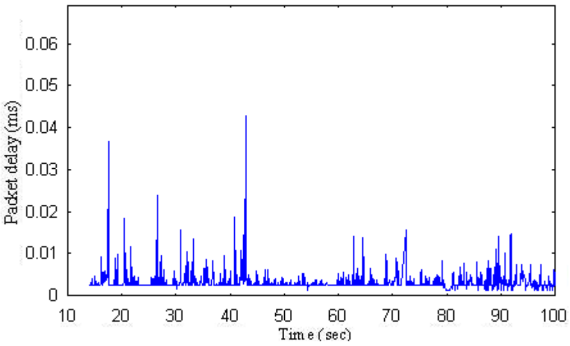


Figure 10: The packets delay in FuzzyMARS with 30 TCP flows vs. simulation time

The results obtained in this section illustrate that the proposed model can support real-time traffic with low delay in one single channel environment. In what follows, we will investigate the case of a multihop network under various traffic and mobility scenarios.

4.2 Performance in multihop environment

The multihop network considered in this section consists of 50 mobile nodes. The network area has a rectangular shape of 1500m x 300m that minimizes the effect of network partitioning. The flows traverse three intermediate nodes on average between source and destination. The network area has a rectangular shape of 1500m x 300m that minimizes the effect of network partitioning. The AODV protocol [18] is chosen as the routing protocol referred to in Figure 1. In this multihop network, we consider a mixture of real-time and TCP best-effort traffic. The real-time traffic is modeled as 4 voice and 4 video flows. The TCP traffic is modeled as a mixture of web and FTP traffic.

The scalability impact of the increasing number of TCP flows on the average end-to-end delay and throughput of traffic is presented in Figures 11-14. Figure 11 illustrates a significant difference in terms of the average delay between the proposed model and the original model. The average delay in FuzzyMARS grows slowly with the increasing number of TCP flows, and it remains almost less than 3 msec. In contrast, the average delay in the original model grows from 7 to 31 msec as the number of TCP flows increases from 2 to 12 flows. Hence, the gain achieved by the proposed model, in terms of the average end-to-end delay, is by about 70-85%. Figure 12 shows the average end-to-end delay in both FuzzyMARS and SWAN models. It is shown that the average delay remains almost less than 3 msec in the proposed model, whereas in SWAN model the average delay is around 5 msec. This means that the achieved gain is about 41% in terms of average delay. We observe the impact of growing number of TCP flows on the average throughput in Figures 13 and 14. The average throughput of the TCP traffic in the proposed model is about 25% less than the original model. This difference is less noticed between FuzzyMARS and SWAN as shown in Figure 14. The previous results confirm that by adopting the FuzzyMARS mechanisms, we can achieve a reduction in the average end-to-end delay by about 70-85% in comparison to the original model at a cost of about 25% decrease in throughput. In addition, the average delay in the proposed model remains almost below 3 msec while the average delay in SWAN grows above 5 msec.

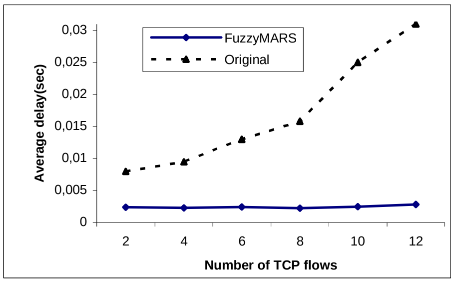


Figure 11: Average delay in the original and FuzzyMARS models vs. number of TCP flows

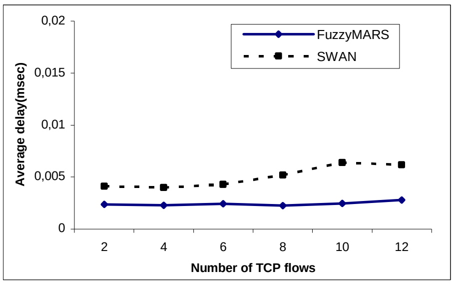


Figure 12: Average delay in FuzzyMARS and SWAN models vs. number of TCP flows

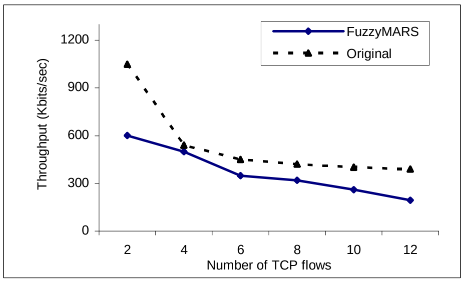


Figure 13: Average throughput in the original and FuzzyMARS models vs. number of TCP flows

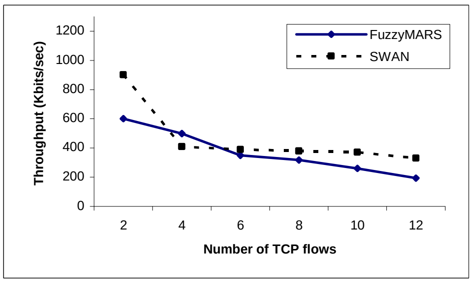


Figure 14: Average throughput in FuzzyMARS and SWAN models vs. number of TCP flows

Figures 15-18 investigate the impact of mobility on the performances of FuzzyMARS. For that end, the random waypoint mobility model [19] is implemented at each node in the network. In the beginning, the nodes are randomly placed in the area. Then, each mobile node selects a random destination and moves with a random speed up to a maximum speed of 20m/s. After reaching the destination, the node will stay there for a given "pause time". When the pause timer expires, the mobile node picks another random destination and moves at another random speed. The mobility scenarios consist of 4 video flows, 4 audio flows, and 10 TCP flows. The real-time traffic is modeled as previously discussed. The best-effort TCP traffic consists of 5 web flows and 5 FTP flows.

We observe in Figure 15 that the average end-to-end delay in FuzzyMARS increases slowly and it grows only for the highest mobility scenarios. The average delay in the proposed model remains almost less than 6 msec, whereas the average delay in the original model grows from 25 to 38 msec. This means that the proposed model achieves a reduction in terms of average delay by about 80-85%. On the other hand, it is observed in the Figure 17 that the throughput of TCP best-effort traffic decreases slowly in both the original and FuzzyMARS models as the mobility increases. This decrease in terms of throughput by about 12-25% when the mobility becomes high is due essentially to the congestion and the broken links in the route relaying source and destination.

Figure 16 illustrates the average end-to-end delay with different mobility scenarios in both FuzzyMARS and SWAN models. It is observed that the average end-to-end delay of traffic in FuzzyMARS increases slowly as mobility increases, and it grows only for the highest mobility scenarios. For different mobility scenarios, the average delay offered by the proposed model is about 15-33% better than that offered by SWAN. On the other hand, it is shown in Figure 18 that for different mobility scenarios, the throughput in SWAN is slightly better than in FuzzyMARS model. SWAN acts better by about 14% better than the proposed model. During the mobility of nodes, some flows are dropped in both SWAN and FuzzyMARS models because of the difficulty in capturing the dynamics of the environment in the ad hoc network.

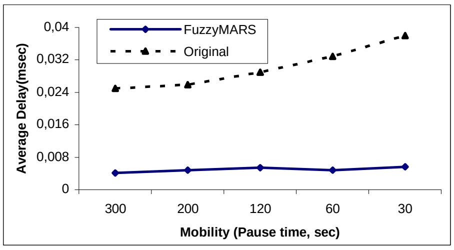


Figure 15: Average delay in the original and FuzzyMARS models vs. mobility

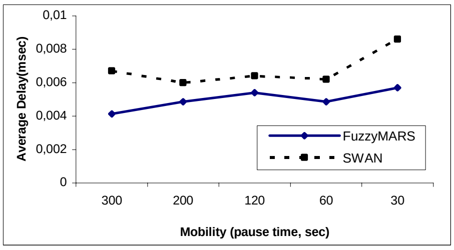


Figure 16: Average delay in FuzzyMARS and SWAN models vs. mobility

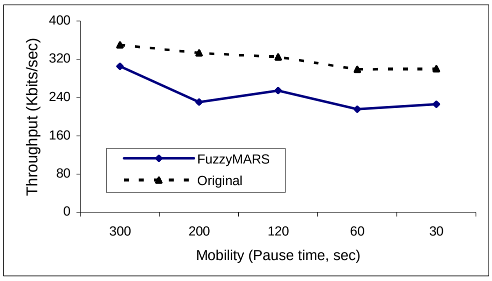


Figure 17: Average throughput in the original and FuzzyMARS models vs. mobility

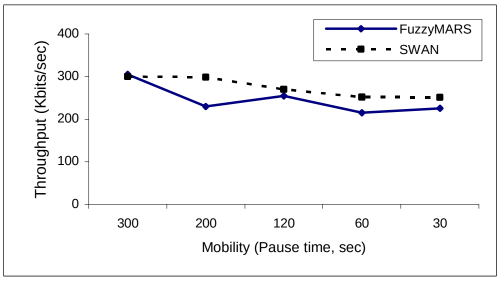


Figure 18: Average throughput in FuzzyMARS and SWAN models vs. mobility

Figures 19-21 show some trace graphs that illustrate the impact of nodes scalability on the average delay in FuzzyMARS in multihop environment. The simulation consists of a mixture of 8 real-time flows and TCP flows, which are modeled as in the previous simulations. These Figures trace the packet delays versus the simulation time using respectively, 10, 20, and 50 nodes. It is observed that the average end-to-end delay in the proposed model increases as the number of nodes becomes high. For different scenarios of nodes scalability in the proposed model, the average delay grows slowly, and the traffic delay becomes almost constant with small and medium number of nodes, and it grows only for the highest scalability scenarios. Figures 19, 20 and 21 show an average delay of respectively, 2.08 msec, 2.26 msec, and 2.56 msec. This means that, even at highest nodes scalability, the proposed model can support real-time traffic with low delay.

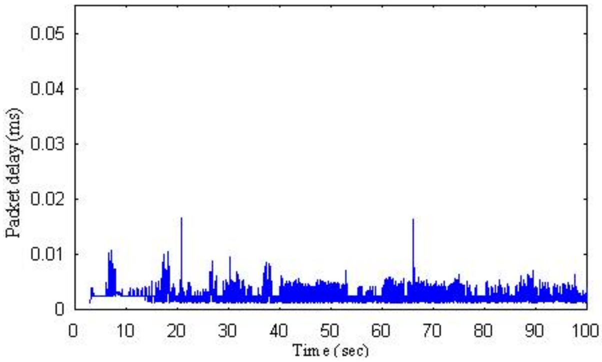


Figure 19: The packets delay in FuzzyMARS with 10 nodes vs. simulation time

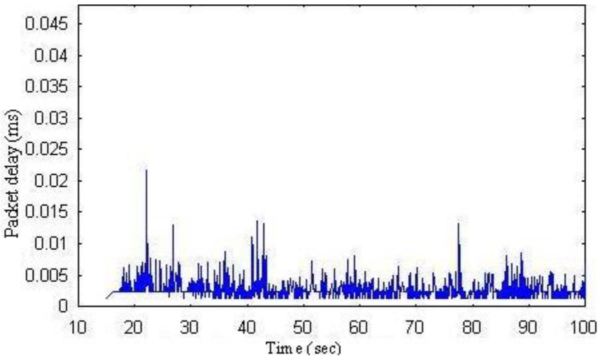


Figure 20: The packets delay in FuzzyMARS with 20 nodes vs. simulation time

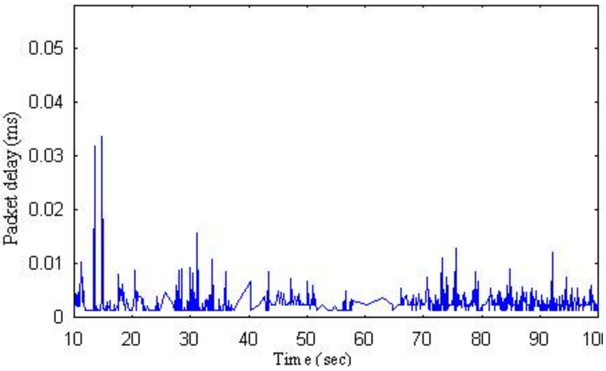


Figure 21: The packets delay in FuzzyMARS with 50 nodes vs. simulation time

5. Conclusion

In this paper, we presented a semi-stateless QoS approach based on a fuzzy logic system for service differentiation delivery in wireless mobile ad hoc networks. The proposed model, named FuzzyMARS, explores how fuzzy logic, for traffic regulation, used with a semi-stateless approach can provide better performances than other counterparts. FuzzyMARS is composed essentially of an admission controller mechanism, techniques for real-time traffic regulation, and fuzzy logic system for best-effort traffic regulation. The performance evaluation of the proposed model was thoroughly studied with ns-2 under different mobility, channel, and traffic conditions. Simulations show the benefits of the proposed over both IEEE 802.11 MAC and SWAN models. It is observed that FuzzyMARS experiences low and stable delays under different channel conditions, traffic scalability, and network mobility while sensibly preserving the throughput. The performance results show that fuzzy logic promises to offer an efficient means for overcoming QoS delivery fluctuations in ad hoc networks, and deserves further attention and study.

Acknowledgement

This work has been performed with the support of the National Sciences and Engineering Research Council of Canada.

References

[1] S.-B. Lee, G.-S. Ahn, X. Zhang, and A.T. Campbell, “INSIGNIA: An IP-Based Quality of Service Framework for Mobile Ad Hoc Networks,” *Journal of Parallel and Distributed Computing (Academic Press), special issue on wireless and mobile computing and communications*, vol. 60, no. 4, pp. 374-406, Apr. 2000.

[2] G.H. Ahn, A. T. Campbell, A. Veres, and L. H. Sun, “SWAN: Service Differentiation in Stateless Wireless Ad Hoc Networks”, *IEEE INFOCOM 2002*.

[3] J.L. Sobrinho and A.S. Krishnakumar, “Quality-of-Service in Ad Hoc Carrier Sense Multiple Access Networks,” *IEEE Journal on Selected Areas in Communication*, vol. 17, no. 8, pp. 1353-1368, Aug. 1999.

[4] L. Khoukhi, S. Cherkaoui, “FuzzyMARS: A Fuzzy Logic Approach with Service Differentiation for Wireless Ad hoc Networks”, IEEE WirelessCom, The International Conference on Wireless Networks, Communications, and Mobile Computing, June 2005.

[5] L. Khoukhi, S. Cherkaoui, “A Quality of Service Approach Based on Neural Networks for Mobile Ad hoc Networks”, *IEEE- IFIP International Conference on Wireless and Optical Communications Networks WOCN 2005*, Dubai, UAE, March 6 - 8, 2005.

[6] H. Xiao, W. K.G. Seah, A. Lo, and K. Chaing, “Flexible QoS Model for Mobile Ad-hoc Networks”, *IEEE Vehicular Technology Conference*, vol. 1, pp 445-449, Tokyo, May 2000.

[7] L. Khoukhi, S. Cherkaoui, Flexible QoS Routing Protocol for Mobile Ad Hoc Networks, In Proc. of the 11th IEEE International Conference on Telecommunication (ICT2004), Brazil, Aug. 2004.

[8] C. R. Lin and J.-S. Liu., QoS Routing in Ad Hoc Wireless Networks, IEEE Journal on Selected Areas in Communication, Vol. 17, No. 8, 1426–1438, 1999.

[9] Qi Xue and Aura Ganz., Ad hoc QoS on-demand routing (AQOR) in mobile ad hoc networks, Journal of Parallel and Distributed Computing, Elsevier Science, USA, 2003.

[10] A. Munaretto, H. Badis, K. Al Agha and G. Pujolle, A Link-state QoS Routing Protocol for Ad Hoc Networks”, In the proceedings of IEEE MWCN02, Stockholm, Sept.2002.

[11] S. Chen and K. Nahrstedt., Distributed Quality-of-Service in Ad Hoc Networks, IEEE Journal on Selected Areas in Communications, vol.17, no. 8, Aug., 1999.

[12] W.-H. Liao, Y.-C. Tseng, J.-P. Sheu, and S.-L. Wang., A Multi-Path QoS Routing Protocol in a Wireless Mobile Ad Hoc Network, In Proc. of IEEE ICN’01, International Conference on Networking, Part II, pp. 158–167, July, 2001.

[13] R. Sivakumar, et al., CEDAR: a core extraction distributed ad hoc routing algorithm, IEEE Journal on Selected Areas in Communications, vol. 17, no. 8, pp. 1454-1465, 1999.

[14] S. Chen and K. Nahrstedt, On finding multi-constrained paths, IEEE International Conference on Communication, 874 -879 June 98.

[15] Z. Wang and J. Crowcroft, QoS routing for supporting resource reservation, IEEE Journal on Selected Areas in Communications, vol. 14, no. 7, 1996.

[16] C.-R. Lin., On-Demand QoS Routing in Multihop Mobile Networks, In Proc. of IEEE INFOCOM 2001, pp. 1735–1744, April 2001.

[17] Y. Chen, Y. Tseng, J. Sheu, and P. Kuo, On-Demand, Link-State, Multi-Path QoS Routing in a Wireless Mobile Ad-Hoc Network, European wireless, Florence, Feb. 2002.

[18] C.E. Perkins, E.M. Royer, Ad-hoc on-demand distance vector routing, In the Proc. of the Second IEEE Workshop on Mobile Computing Systems and Applications, pp.90 – 100. New Orleans, LA, Feb. 1999.

[19] J. Broch, D. Maltz, D. Johnson, Y.-C. Hu, and J. Jetcheva, “A Performance Comparison of Multihop Wireless Ad Hoc Network Routing Protocols”, ACM/IEEE Mobile Computing and Networking, Octobre 1998.

[20] X. Wang, “A Course in Fuzzy Systems and Control”, *Upper Saddle River*, NJ: Prentice-Hall, 1997.

[21] C.C Lee, “Fuzzy logic in control systems: fuzzy logic controller- part I and II,” *IEEE transactions on systems, man, and cybernetics*, vol. 20, no. 2, 1990.

[22] L. A. Zadeh, “fuzzy logic = computing with words”, *IEEE Transactions on fuzzy system*s, vol. 4, no2, pp. 104-111, 1996.

[23] T. Takagi and M. Sugeno, “Fuzzy identification of systems and its applications to modeling and control,” *IEEE transactions on systems, man, and cybernetics*, vol. SMC-15, pp. 116–132, Feb. 1985.

Dr. Lyes Khoukhi is a Post doctoral researcher at the University of Montreal (Canada) since 2008. He received Ph.D degree in Electrical and Computer Engineering from the University of Sherbrooke (Canada) in 2007, and M.Sc degree in Computer Engineering from University of Versailles (France) in 2002. During 2003-2007, he stayed in INTERLAB Communications Research Laboratory, Sherbrooke University. His research interests include wireless communications, mobile ad hoc networking, intelligent systems, and distributed systems.

Dr. Soumaya Cherkaoui is an Associate Professor at Sherbrooke University, Canada which she joined in 1999. Since 2000, she is also the director of INTERLAB, a research group comprising more than 12 faculty and research assistants which conducts research funded both by government and industry. Before joining Sherbrooke University as a faculty member, Professor Cherkaoui worked for industry as a project leader. She has over 50 publications in the areas of network protocols and distributed systems.