The Full ESWAN Destination-Based Approach:
Operations And Evaluation

Yasser L. Morgan[1], Thomas Kunz[2]

[1]*School of Information Technology*
[2]*Systems and Computer Engineering,*
Carleton University

Abstract: In response to the growing need to support better than best-effort (BE) quality of service (QoS) in mobile ad-hoc and sensor networks, many QoS models have been proposed. SWAN independent QoS model is introduced to operate on wireless ad-hoc networks. As a cross layer QoS model, SWAN is flexible and may run over any routing protocol or Media Access Control (MAC) layers. SWAN provides some advantages over competitive models However, SWAN is vulnerable to problems related to mobility and false admission. The original SWAN model discusses the two problems as part of a dynamic regulation of real-time flows, and introduced two solutions, namely source and network-based regulation algorithms. This paper criticizes both regulation algorithms and show why destination-based algorithm selects real-time victim flows in a better way. Then we provide test results to analyze and evaluate the destination-based approach.

Index Terms: QoS, Ad-hoc Networks, Mobility

I. Introduction

QoS support in ad-hoc and sensor networks has become an active research topic. Recognizing the fact that QoS frameworks for the fixed Internet will not be appropriate for networks with highly dynamic topologies, researchers have proposed a number of QoS solutions for ad hoc networks. Among those proposed, SWAN [1] has illustrated higher levels of robustness and ability to recover from adverse mobility situations when compared to FQMM [17], dQoS [10], or INSIGNIA [9], due to its stateless model. SWAN can operate over Best Effort (BE) Media Access Control (MAC) such as IEEE 802.11 Distributed Coordination Function (DCF) [20], and uses a stateless distributed approach to solve the dynamic QoS issues. SWAN operates in a fully decentralized manner in order to deal with the ad-hoc network dynamics. SWAN

uses source admission control to limit the amount of admitted real-time flows. In response to network dynamics, which leads to occasional congestion, SWAN uses explicit congestion notification (ECN) to dynamically regulate real-time traffic. Since intermediate nodes do not maintain per-flow state information, solving congestion scenarios becomes a bit challenging. However, maintaining the tenet of stateless model keeps the system simple, robust, and scalable.

SWAN also adapts the concept of "soft" real-time service guarantees [5]. When a real-time flow is admitted to the network using admission control, it is possible at any point during the lifetime of this flow to be downgraded to best-effort, or to stop it in response to network dynamics. A source node will have to re-initiate a new admission process in order to, possibly, re-establish the flow. This is a powerful characteristic for SWAN since the soft real-time guarantee acts as a formal response to real situations that ad-hoc wireless networks may experience.

However, SWAN implements no service differentiation based on user profiles as in [18]. We feel that SWAN approach in this regard is more consistent with the nature of ad-hoc and sensor networks in the sense that those network form and disappear rapidly with little chance to configure and set user profiles. User equipments are usually of low computational and storage capabilities.

SWAN also has ignored the dimension of real-time flow delays. The SWAN framework uses per-link delays merely to detect congestion, and uses bandwidth estimates to facilitate admission control. End-to-end delays have not been considered in measuring the quality of real-time flows in SWAN. When source and destination nodes are away from each other (in terms of number of hops), real-time packets experience higher end-to-end delays. When end-to-end delay of a real-time flow approaches a certain threshold, the flow becomes highly sensitive to network dynamics. We argue that destination nodes can make better judgment on the

quality of received real-time flows by using end-to-end delays. Therefore, we propose a destination-based algorithm to solve the dynamic regulation issues.

In this paper, we illustrate briefly the relevant elements of the original SWAN model in Section 2. Then, in Section 3, we describe the problem of dynamic regulation of real-time flows. In Section 4 we show the two proposals provided within the original SWAN framework, namely source-based and network-based algorithms, then criticize both proposals. In Section 5 we introduce some terminologies essential to discussing the dynamic regulation and introduce the destination-based algorithm. In addition, we illustrate the rational behind our destination-based algorithm, and the reason it selects a better set of victim flows. In Section 6 we describe the test-bed used to validate the destination-based algorithm, and illustrate the test-bed, the results, analyze, and evaluate the proposed enhancements. Finally in Section 7 we conclude the research and suggest possible future work.

II. Common SWAN Operations

SWAN can be described in terms of two major components. The *Admission Controller (AC)* is responsible for admitting any new flow to the ad-hoc network. Admission control is performed at the source node that initiates real-time flows. The *Rate Controller (RC)* is responsible for regulating (BE) traffic and maintaining traffic loads at intermediate nodes. Other components are a *Classifier*, which selects real-time packets to bypass the shaper, and a *Shaper*, which represent a simple leaky bucket traffic shaper. The goal of the shaper is to delay BE packets in conformance with the rate calculated by the *Rate Controller*. Figure 1 illustrates the architecture of SWAN framework.

II.1 Admission Controller (AC)

In SWAN, a mobile ad-hoc node can initiate a real-time flow only after obtaining an accept message. SWAN-AC adopts the well-known Additive Increase Multiple Decrease algorithm (AIMD), which has been used by the TCP congestion controller for decades. The TCP congestion control algorithm ensures that the system works around, or preferably close to the "cliff" as illustrated on Figure 2, which ensures maximum system throughput at the expense of having larger queues, and therefore, longer average delays [6].

The SWAN AIMD admission controller algorithm follows a comparatively conservative approach, SWAN keeps the system at the delay "knee",

where the system throughput is almost the same as at the cliff, but queues are significantly less loaded [6]. SWAN uses MAC delay as a feedback instead of packet loss since losses typically happen at the cliff while delays develop at the knee. This approach is illustrated in [15], [16], and fully described in [1], [2].
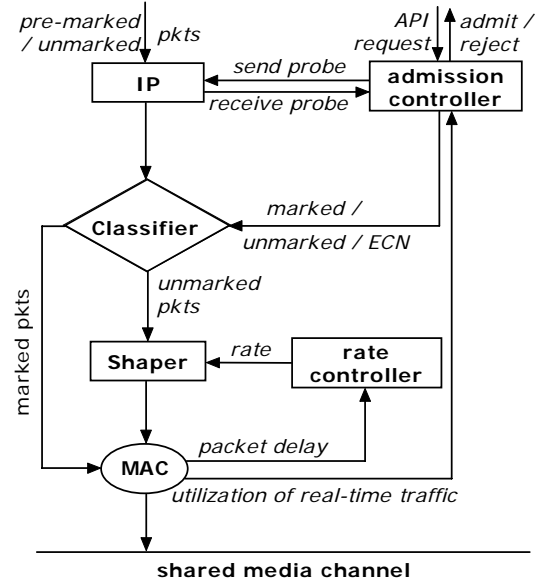

Figure 1: *Block Diagram of the SWAN Architecture*

As a result of this conservative approach, SWAN uses for real-time flows less than the available bandwidth, assuming that the remaining bandwidth will be consumed by (BE) traffic. In addition, this slack of bandwidth can be viewed as a safety measure against the ad-hoc dynamics such as bandwidth variation and mobility.
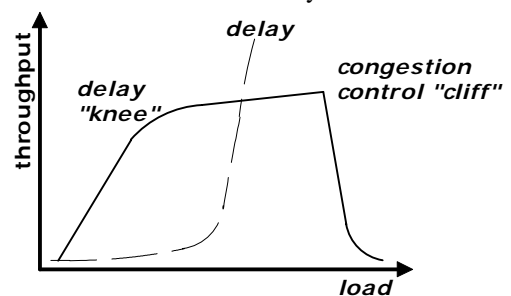

Figure 2: *General Behavior of a Congestion Controlled System*

AC starts by sending a probe request to the destination node. Intermediate nodes intercept the request and update it with the bottleneck bandwidth. Intermediate nodes use their AC to calculate available bandwidth for newer real-time flows, but they do not implement resource, or bandwidth allocation. The destination node replies with a probe response that contains the bottleneck bandwidth. Once the source node receives the

probing response packet, it can execute the source-based admission control by comparing end-to-end bandwidth availability with the bandwidth requirement for the new real-time flow.

II.2 Classifier

The source node marks packets associated with admitted real-time flow as real-time (RT for short). The classifier sends marked packets to the MAC layer directly, bypassing the shaper. SWAN implicitly assumes that real-time flows need not to be policed.

II.3 Shaper

The shaper is a simple leaky delay queue that enforces delays on (BE) packets based on feedback from the RC.

II.4 Rate Controller (RC)

The RC determines the link status using link delay measured by the MAC layer. The delay, for instance, can be extracted using the IEEE 802.11 DCF mechanism. The RC detects excessive link delay when one or more packets have greater delays than a threshold link delay d (sec). The threshold delay d is based on the real-time delay requirements as discussed in [15], [16]. When excessive link delay is detected, RC backs off the rate by $r\%$ (multiplicative decrease by $r\%$). RC has to re-adjust parameters (d and r) every T seconds. The diagram shown in Figure 2 and most of this section refer to [1], [2], and [3].

III. Dynamic Regulation Issues

SWAN introduces dynamic regulation mechanisms in response to conditions raised by network dynamics like node mobility, and false admission. It is important to illustrate the impact of both issues on network resources.

III.1 Mobility

As illustrated in Figure 3, real-time flows between nodes s and d can be redirected from node n_1 to node n_2 due to mobility, and the underlying routing algorithm will perform the necessary rerouting. However, node n_2 will experience an increase in real-time traffic even though it did not perform any admission process to allow the new flow. This situation is referred to as congestion due to mobility.
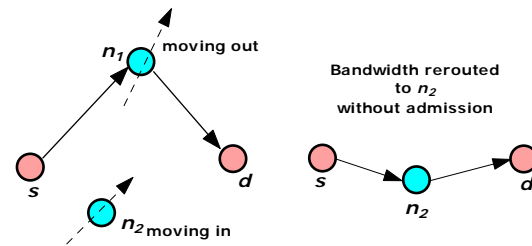

Figure 3: *Congestion/Overload Due to Mobility*

III.2 False Admission

As illustrated in Figure 4, nodes s_1, s_2, and s_3 may initiate a probe request to send real-time flows to nodes d_1, d_2, and d_3 (respectively) through node n. If node n processed the three requests within a short time (i.e. before real-time packets start arriving at n), the admission controller at node n will accept the three flows even if it practically has room for only one flow. This is due to the lack of resource reservation in SWAN. Until real-time packets consume available bandwidth, node n will always admit new real-time flows. This situation is referred to as congestion due to false admission.
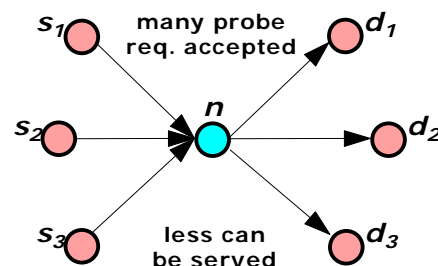

Figure 4: *Congestion/Overload Due to False Admission*

It is important to realize that mobility and false admission merely represent two issues among other issues related to network dynamics. For instance, deterioration in radio link quality may occur due to interference, introduction of a barrier, or due to diminishing battery life. The term network dynamics commonly refers to issues related to mobility, radio quality, and distributed operation. SWAN adopts the explicit congestion notification (ECN) regulation algorithm to recover from congestion conditions caused by network dynamics. Since nodes are continuously (and independently) monitoring their bandwidth utilization, nodes can detect violations. Congested nodes will then use the ECN bits in the IP header of the real-time packets to inform destination(s) of the existence of congestion. Each destination node will issue a regulate message to the relevant source node. Source nodes will then re-initiate new probe requests to locate possibly a better service route to

the destination, or terminate the flow due to lack of resources.

IV. Common Dynamic Regulations

The decision of congested nodes to mark packets with ECN is very critical because flows that get marked with ECN may lose their QoS privileges. SWAN has proposed two regulations, namely source and network-based regulations. Both approaches mark ECN packets differently, but use the same consequences afterwards.

1. **Source-Based Regulation:** In source-based regulation, a congested node marks all RT flows with (congestion experienced) CE using the ECN bits. When destination nodes encounter packets with CE bit marked, they send a regulate message to the associated source nodes. Source nodes immediately perform multiplicative decrease on relevant RT flows. As a result, the congested node experiences a gradual decrease in the amount of RT traffic until the congestion condition is removed, at that point, the intermediate (previously congested) node stops marking CE bits. If the used bandwidth of a specific RT flow is unsatisfactory to its source node, it has to backup a random amount of time then re-initiate a probe request to re-establish the desired level of service, possibly on a different route. Source nodes have to stagger the re-initiation in order to avoid a flash-crowd condition where nodes may fall into another false admission again, therefore, the random backup time is essential. A source node may perform a biased re-initiate flow towards newly admitted real-time flows if it can keep information about newly admitted flows.

   Source-based regulation forces all RT flows going through one congested node to be regulated. This approach seems to be too aggressive. It can force too many flows to be regulated even if the amount of bandwidth violation is limited. In addition, it does not discriminate between different RT flows. Following that approach, some limited-quality RT flows might maintain connectivity, while other higher-quality RT flows might be denied service re-initiation or un-necessarily get disrupted.

2. **Network-Based Regulation:** In a network-based regulation, a congested node selects a subset of all its real-time flows to be a "congestion set" or "victim flow set". The congested node marks packets associated with victim flows only. It is possible for a congested node to distinguish a specific set of

RT flows by applying a simple hash function without any need to keep flow information. Packets of victim RT flows will reach relevant destination nodes marked with CE then the network-based approach follows the same process as described for source-based regulation. If a congested node does not experience any decrease in the amount of real-time traffic after a period of time T seconds, it calculates a new set of victim flows. SWAN suggests applying some intelligence at the congested node in order to select the set of victim flows. For instance, if source nodes inject RT-flows with RT-old or RT-new, using the IP-TOS field, congested nodes can use a biased function to form the set of victim flows out of newer flows, hoping to decrease false admission.

   Network-based regulation selects the victim flows set randomly, and in the best case, discriminates against newly admitted flows. However, the idea of loading the IP-TOS field may conflict with flows that need to use this field, especially when the flow extends over the Internet as in [11].

V. Destination-based Regulation

In the following subsections, we compare the proposed destination-based regulation to other destination-based mechanisms and show the difference. Following we identify the proposed approach and elaborate on its preemptive and recovery behaviors.

V.1 Common Destination-based Approaches

The term destination-based approach is commonly used by a breed of QoS routing algorithms in an attempt to enhance the classic per-source-destination QoS routing granularity. Perhaps the most obvious destination-based QoS routing algorithm is in [24] where QoS flows are marked on a per-destination granularity for simplicity in a compromise to lower the algorithm complexity at intermediate nodes. Other destination-based QoS routing algorithms used per-destination granularity combined with maintaining a State-Path as in [22] or combined with hybrid Genetic Algorithm as in [23]. In all destination-based QoS routing algorithms, QoS flows are identified on a per-destination granularity in an attempt to mimic ATM solutions. The classical use of destination-based approaches can be broaden to serve other objectives like in [21] where the approach is used to scale bandwidth brokerage and service provisioning.

The destination-based approach proposed in this research is quite different. Most QoS mechanisms rely on the source to identify the per-flow QoS parameters, reserve the end-to-end resources, and then transmit. Our proposal uses feed back information from destination to help intermediate, congested, nodes identify which QoS flows are suffering from network mobility, and hence, can be removed to release the congestion. Therefore, the destination-based approach presented here is fairly unique.

V.2   Basic Definitions

The proposed enhanced SWAN with destination-based regulation (ESWAN for short) evolved from our findings about SWAN behavior in response to above-average traffic load. By above-average traffic load we mean at least one third of available bandwidth is consumed by real-time traffic and one third by best-effort traffic. The delay histogram shown on Figure 5 illustrates that the majority of RT packets experience a delay of less than 35 msec in the above-average traffic load condition.
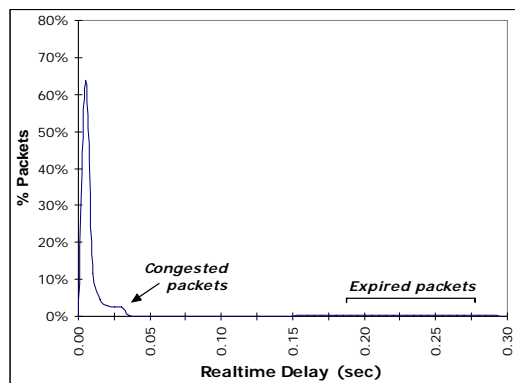

Figure 5: *SWAN RT packet delay histogram*

However, a considerable percentage of packets appear to experience delay higher than 175 msec. This is clear on the cumulative graph in Figure 6, where about 9% of RT packets seem to experience delays beyond the 175 msec. Interactive VoIP flows, for instance, will disregard packets with delays beyond a certain threshold (150 msec) (called expired packets for short). When repeating the same test for various mobility scenarios, the SWAN model consistently caused about 9% of the RT packets to expire (9.04%, 9.64%, 9.97%, 9.72%, 10.03%, and 9.87%). Therefore, the bandwidth consumed by highly delayed packets (expired bandwidth for short) is a bandwidth that unnecessarily consumes network resources, and degrades services provided to other RT flows. Optimizing network resource utilization requires decreasing the amount of expired bandwidth.

Destination-based regulation relies on destination nodes to detect an increase in expired bandwidth, then, regulate each flow accordingly. This preemptive behavior is followed to prevent wasting network resources. In case of congestion, the destination-based regulation selects a subset of the congested flows in order to regulate. This subset is selected based on flow quality starting with the lowest quality flows first. The quality of each flow is measured as a function of its packet delays.
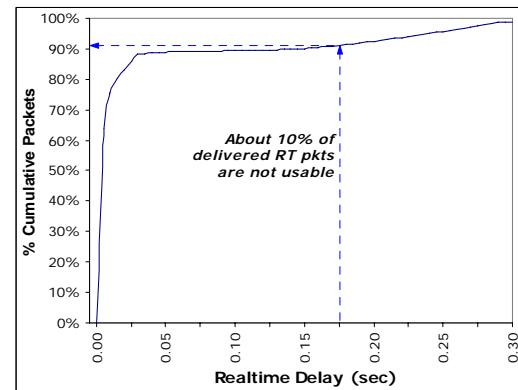

Figure 6: *Cumulative RT packet delay %*

Maximum acceptable packet delay (MAPD) is the threshold packet delay value (in sec) that will certainly result in the destination application ignoring packets of a specific flow. Therefore, MAPD is a flow specific value, known to the destination node, and can be compared to the end-to-end packet delay.

Expired packets are real-time flow packets that exhibit a delay more than MAPD. In the same way, **expired bandwidth** is the bandwidth consumed by expired packets. Also *effective bandwidth = received bandwidth – expired bandwidth.* Therefore, **effective bandwidth** is the bandwidth realized at a destination node and is usable by destination application to replay the real-time flow over a period of time *T*.

Limited QoS is the QoS a network provides to a flow where the required bandwidth is more or less delivered, but a significant part of the received bandwidth is not usable due to excessive packet delays. The effective bandwidth perceived by a destination node over a period of time *T* is hardly sufficient for the application to effectively replay the real-time flow.

Effective bandwidth ratio (EBR β) is the percentage of effective to received real-time bandwidth at a destination node for a specific real-time flow over a period of time *T*. Since the bandwidth ratio (β) is based on effective bandwidth then, $0 \leq \beta \leq 1$. Figure 7 and equation 1 illustrate the (EBR β), and limited QoS definitions.

$$EBR \ \beta = \frac{effective \ BW}{received \ BW}\bigg|over \ time \ T \qquad \dots (1)$$

EBR (β) measures the quality of a RT flow where values closer to *1* indicate high flow quality, and values closer to *0* indicate limited flow quality and inefficient bandwidth usage. The RT flow-specific EBR values (β_H, β_L) represent desired measures of real-time flow quality, where values lower than β_L represent a waste in network resources that require regulation. The probe request message is used to communicate both β_H, and β_L values to the destination node.
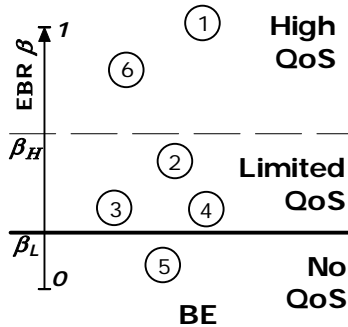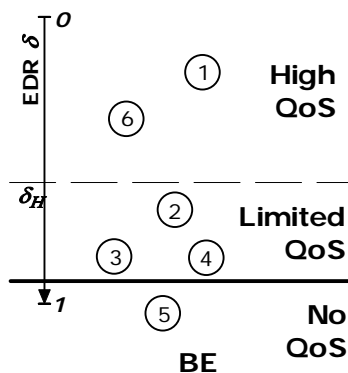

Figure 7: *(EBR β) Service view in a loaded inter-mediate node*

Effective delay ratio (EDR δ) is the percentage of average effective packet delays at a destination node to MAPD, over a period of time *T*. Since delay ratio (δ) is based on effective bandwidth only, $0 \leq \delta \leq 1$.

$$EDR \ \delta = \frac{avg. \ eff. \ pkt \ delay}{MAPD}\bigg|over \ time \ T \qquad \dots (2)$$

Effective delay ratio measures the quality of the flow. EDR values closer to *0* indicate higher QoS, while higher EDR values closer to *1* indicate a flow that is suffering from high delay averages, but the quality is still acceptable, as described in equation 2.

Figure 8: *(EDR δ) Service view in a loaded inter-mediate node*

As illustrated in Figure 7 and Figure 8, a real-time flow like flow 1 (denoted by circle 1) is a flow that is getting better than required bandwidth, and

therefore has high QoS. Flows 2, 3, and 4 are having limited QoS since they are getting the required bandwidth, but their effective bandwidth is hardly at the required limit.

Flow 5 is a flow demoted to non-QoS (i.e. best-effort), and is not seen by intermediate node as a distinct RT-flow any more. Also flow 1 has a better quality than flow 6, and the difference can be expressed by the flow values of δ and β.

V.3 Preemptive Behavior

The destination-based algorithm is based on two behaviors, the first is a preemptive behavior, which monitors the quality of service provided by the network, and request service upgrades if the provided service is unsatisfactory. The second is a recovery behavior, which is executed when intermediate nodes experience congestion, by regulating limited QoS flows before regulating higher QoS flows.

Destination nodes perform preemptive behavior on flows experiencing limited QoS without detecting congestion condition. Simply, if a sufficient number of packets arrive at a destination node with packet delays higher than *MAPD*, over a period of time *T* the destination node detects a limited QoS condition ($\beta < \beta_L$), and issues a regulate message to the relevant source node. The source node then triggers a re-initiate procedure to locate, possibly, another route with better quality.

V.4 Recovery Behavior

Recovery behavior is also performed by a destination node and implements the following mechanism:

a) If an intermediate node is experiencing congestion, it marks all RT packets with CE using the ECN bits. The marking of packets will continue until the intermediate node realizes a sufficient decrease in the arriving bandwidth.

b) Destination nodes with ($\delta \geq \delta_l$) will have to issue a regulate message immediately.

c) Other destination nodes will wait for time *T*, if packets keep arriving marked with CE, then destination nodes with ($\delta \geq \delta_2$) will have to issue a regulate message immediately.

d) Over time, congestion gets resolved by removing flows with higher relative delays δ first. Values of (δ_i) are constants for the network, and have to be selected such that ($\delta_i > \delta_{i+1}$).

This mechanism enables different destination nodes to, independently; regulate related real-time flows by regulating lower quality flows first. If

congestion is not resolved, flows having slightly better quality are regulated, until congestion is resolved.

VI. Evaluation and Analyses

The initial evaluation view may seem to compare the proposed destination-based approach with both source-based and network-based regulations. However, the source-based regulation is merely a special case of network-based regulation where the subset of victim flows is the global set. Interested readers can review [3] which include a comparison between source and network-based regulations. Our testing showed that aggressive and dramatic behavior of source-based regulation can be easily improved by adopting network-based regulation. Hence this section compares destination-based regulation with the more mature network-based regulation.

VI.1 Test-bed Description

In order to test the destination-based approach, we used the ns-2 simulator [4], [7], and [19]. The test-bed assumes a square field with 20 nodes moving with a max speed of 10 m/s, and pause time of 2 sec. The test-bed uses parameters from a Lucent WaveLAN card to set up a radio link using 802.11 as MAC layer where every mobile node has a transmission range of 250 m. For routing, we use the Ad-hoc On-demand Distance Vector (AODV) [13] and the Dynamic Source Routing (DSR) [8] protocols.

The traffic generator uses both constant and variable bit rate RT-applications (CBR/VBR). In addition, we use TCP connections by simulating greedy FTP applications with packet size 512 bytes. TCP connections provide (BE) packets that do not require QoS services. Real-time VoIP flows establish MAPD of 150 msec, while burst video flows use 450 msec. Finally, the total simulation time is 300 sec, EBR (β_H, β_L) values are (0.97, 0.95) respectively, and EDR (δ) set values are {0.9, 0.8, ..., 0.2, 0.1} for all RT flows.
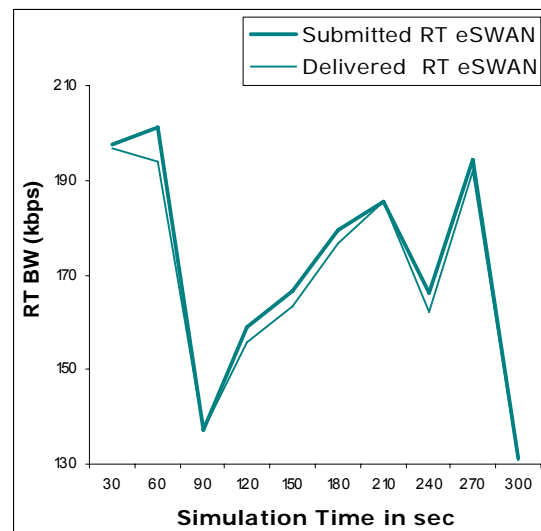
VI.2 Bandwidth Efficiency

In order to evaluate the destination-based approach, both the SWAN and ESWAN test-bed are subject to the same traffic patterns and mobility scenario. The amount of submitted and delivered RT bandwidth is monitored throughout the simulation time. The results are illustrated in Figure 9 for SWAN and Figure 10 for ESWAN.

From Figure 9 & 10, it is easy to see that while the submitted bandwidth are very similar in both SWAN and ESWAN test-bed, the delivered bandwidth has a closer match to the submitted bandwidth in ESWAN than SWAN. The limited difference in submitted bandwidth is due to the use of preemptive behavior in ESWAN that re-initiates RT flow probing.
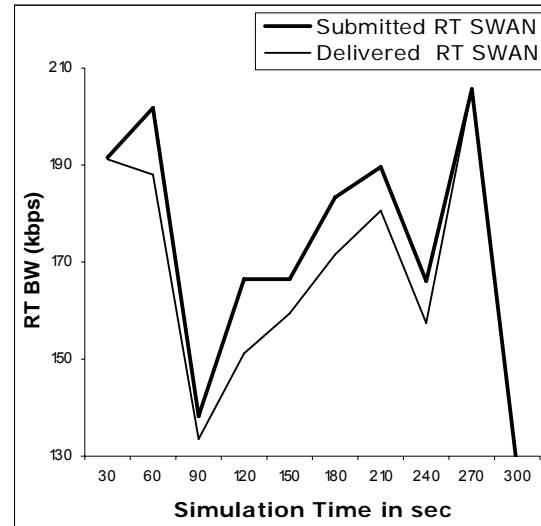
Figure 9: *Efficiency of Bandwidth Usage in SWAN*

ESWAN destination-based approach is focused on the quality and usability of delivered RT bandwidth. In order to consider the effective bandwidth delivered at destination nodes, it is important to consider the expired RT bandwidth and the EBR ratio as defined earlier.


Figure 10: *Efficiency of Bandwidth Usage in ESWAN*

VI.3 Bandwidth Evaluation View

Figure 11 shows the distribution of expired RT bandwidth in both SWAN and ESWAN. The selected traffic pattern triggers the ESWAN
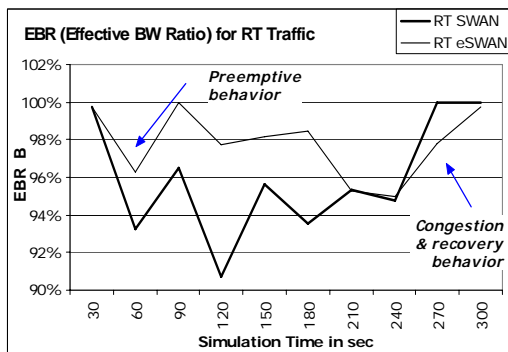
preemptive behavior at about 60 sec after simulation starts, and causes congestion on both SWAN and ESWAN models at about 230 sec from simulation start time.

Running the described test-bed using SWAN and ESWAN models we can observe a considerable decrease in the amount of expired RT bandwidth that consumes valuable network resources (almost one third).
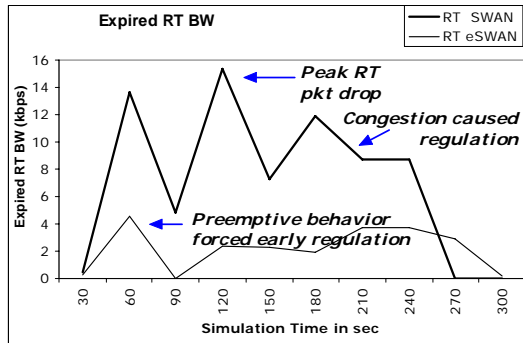
Figure 11: *Distribution of Expired RT Bandwidth*

From Figure 11 it is easy to see that the preemptive behavior causes a virtual ceiling on the amount of expired RT bandwidth. SWAN on the other hand does not impose any limitation on the amount of expired RT bandwidth; instead it regulates flows only after congestion takes place by sensing the link. After about 230 sec of simulation time, congestion occurs and forces the regulation of RT flows in both SWAN and ESWAN simulations. It is interesting to realize that the peak amount of expired RT bandwidth in SWAN is not directly related to congestion as shown on Figure 11.
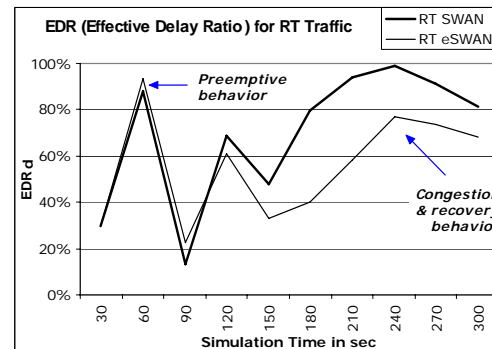

Figure 12: *EBR Distribution for SWAN and ESWAN*

Figure 12 shows the distribution of EBR for both SWAN and ESWAN. The preemptive behavior of ESWAN detects a limited QoS condition at about 60 sec of simulation time, ESWAN forces a re-regulation of relevant RT flows. Therefore, ESWAN shows limited fluctuations of the EBR value. SWAN on the other hand has no limitations on the EBR values; thus, the EBR value goes as low as 90%.

At about 230 sec of simulation time, both SWAN and ESWAN detect congestion, SWAN uses network-based recovery, and ESWAN uses destination-based recovery. It is clear from Figure 12 that ESWAN recovers a little slower, and therefore, causes fewer disturbances to RT flows than the radical recovery of SWAN. It is also important to realize that during all the simulation time before congestion, SWAN did not detect any problem with RT flows, and therefore, did not regulate RT flows. As a result, RT flows on SWAN experience less quality than the network can provide.

VI.4 Delay Evaluation View

In Figure 13, the preemptive behavior triggered at about 60 sec of simulation time forces re-regulation of a RT flow, and therefore, introduces some minor delays in ESWAN compared to SWAN. The congestion, which happens at about 230 sec of simulation time, is quite different. In ESWAN, flows experience, comparatively, a less congested node than SWAN due to the preemptive behavior. The recovery behavior on ESWAN is a little less radical.

Figure 13: *EDR Distribution for SWAN and ESWAN*

It is important to illustrate the effect of the destination-based algorithm on BE packet delays as illustrated on Figure 14. The preemptive behavior which is invoked at about 60 sec of simulation time causes the BE average delays to peak shortly after. The slow recovery behavior of ESWAN causes a slow recovery of BE packet delay as well. Generally, the delay variation of BE traffic is higher for ESWAN than SWAN, however, this may not cause performance issues since BE traffic belongs to elastic applications that are able to tolerate such variations.

Figures 15 & 16 show the histogram, and cumulative distribution of RT packet delay in ESWAN and can be compared to Figures 5 & 6. In ESWAN less than 1.2% of the delivered RT packets were found expired. When repeating the same test for various mobility scenarios, the

ESWAN model consistently caused about 1.2% of RT packets to expire (1.13%, 1.16%, 1.17%, 1.15%, 1.16%, and 1.14%). Destination nodes use this limited percentage to monitor the services provided by the network and force regulation when necessary.
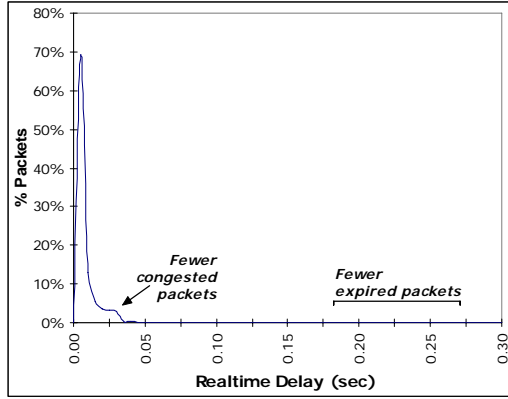

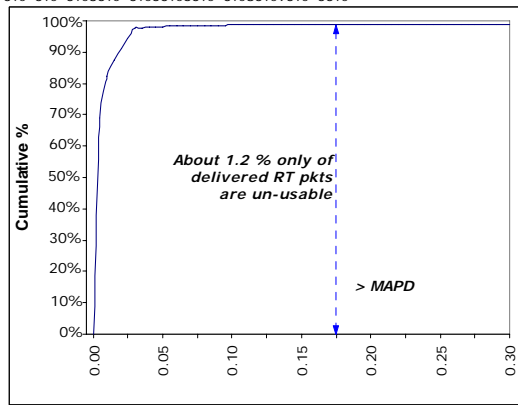**Figure 15:** *RT packet delay histogram in ESWAN*

Figure 16: *Cumulative RT packet delay % in ESWAN*

Comparing the figures observed in this test with the corresponding results from Section 5, the ESWAN model has consistently decreased the percentage of expired RT packets by 7.91%, 8.48%, 8.80%, 8.57%, 8.87%, and 8.73%. In order to calculate the Confidence Interval for the series, we apply equation (3).

$$Confidence\ Interval\ =\ \overline{X} \pm t_{\left(\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}} \ ... \ (3)$$

Where:

$\overline{X}$ = *The mean difference between SWAN & ESWAN observations*
n = *Number of samples (n=6).*
σ = *The standard deviation of the difference between SWAN & ESWAN observations.*
$(1-\alpha)$ = *Confidence level (α=0.05).*
$t_{\left(\frac{\alpha}{2}\right)}$ = *The upper critical value of the **t** distribution (=2.45).*

The confidence interval for the percentage of decrease in expired RT packets when using ESWAN compared to when using SWAN is calculated based on the 6 observation samples. The 95% confidence interval is [8.21%, 8.91%]. As this interval does not include 0, the performance improvement by ESWAN is statistically significant, even with our somewhat limited sample size of 6. This percentage gain in RT packets has a significant impact on the delivered RT quality and on the user perception.

VI.5 The Effect of Mobility

Node mobility is an important factor in the design and evaluation of ad-hoc based technologies. The speed of mobile nodes and their pause time are commonly used attributes to define mobility. The test-bed used pause time of 2 seconds, and when changing the pause time, both SWAN and ESWAN showed little changes in behavior. When running the same test-bed with node speed of {10, ..., 50} meter per second, both SWAN and ESWAN maintained the same level of average packet loss as illustrated in Figure 17 up to node speed of about 35 meter per second.
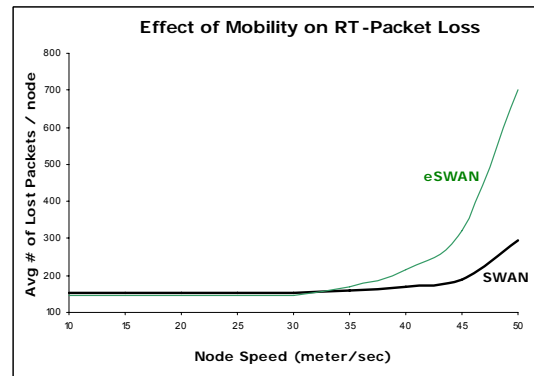

Figure 17: *The effect of node mobility on average packet loss*

When mobile nodes move faster than 35 meter per second, deterioration in radio link quality takes effect. ESWAN shows a higher number of packet losses, and the losses grow much faster compared to SWAN. The reason is the preemptive behavior in ESWAN, which responds to the limited QoS perceived at destination nodes by forcing too many re-initiate probe requests flooding the relevant routes and causing congestion, and packet loss. SWAN, on the other hand, relies on re-routing, which is sufficient in high mobility scenarios.

Therefore, ESWAN is recommended in installations involving limited mobility (i.e. ≤ 35 m/s). We believe this is not a major restriction since the threshold speed here is beyond vehicular speed limits (i.e. ≤ 125 km/hr).

VI.6 Overall Evaluation

In order to investigate the behavior of EBR (β), and EDR (δ) ratios, we apply some changes to the test-bed.

An increasing traffic load is applied to a five mobile nodes test-bed, and the total consumed bandwidth is measured then normalized over a period of time T sec. The mobile nodes are forced to a no mobility condition, and the values of (β and δ) ratios of a VoIP flow are observed against the increasing RT traffic load of the network. Figure 18 and Figure 19 represent the results under these conditions. In both figures, the horizontal axes (average load per node) represent the normalized collective bandwidth consumed by all five nodes for RT flows. Therefore, the exact values of the network RT load will vary based on the test-bed topology, flow directions, setup, and configurations; however, the shape of the curves will remain the same.
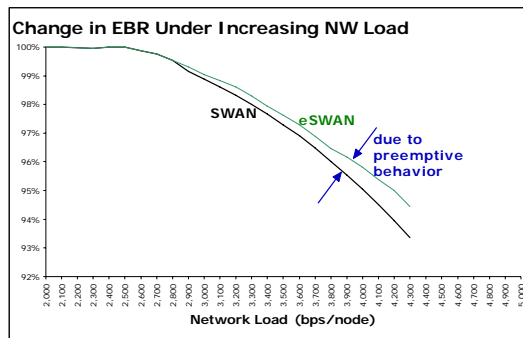
Figure 18: *The effect of network load on EBR*

Figure 18 illustrates the impact of increasing overall network RT load on the EBR (β). Due to the preemptive behavior, ESWAN tends to show higher EBR (β) values than classical SWAN. EBR (β) values lower than 95% are regulated by ESWAN, and the re-initiation of RT flows provides either higher EBR (β) value, or the RT flow will be denied service, and hence, have no EBR (β).

Figure 19 illustrates the impact of increasing overall network RT load on the EDR (δ). When network RT load increases, the effective average packet delay increases, hence the EDR (δ). There are relatively lower values for EDR (δ) on ESWAN than SWAN due to the recovery behavior. High values of δ ($> 70\%$) are commonly associated with congestion, while low values ($< 10\%$) are associated with healthy RT flows.

VII.   Comments and Conclusion

The original SWAN model discusses the source and network regulations algorithms as part of a dynamic regulation of real-time flows, and introduces the two regulations namely to provide full congestion recovery. Both SWAN solutions apply random (or almost random) selection to victim flows, and therefore add little value to the model. This paper introduces the new destination-based regulation to enhance the congestion recovery of real-time flows rather than the source or network based regulations. The destination-based regulation uses a biased rule to select victim flows, and adds a preemptive behavior to decrease the frequent occurrence of congestion.

Packets traveling over larger ad-hoc or sensor networks are likely to experience longer delays since they travel over more hops. Enforcing MAPD threshold, using the EDR (δ), enables the network to limit expired bandwidth, which releases part of the traffic load, and ultimately increases both bandwidth availability and effective use of RT bandwidth. This enhancement comes at the expense of BE traffic that realizes relatively higher average delays, but has only a minor influence on the BE bandwidth. The preemptive behavior is shown to smooth the resource utilization over time, and to decrease chances of congestion. In addition, it enables destination nodes to monitor the actual level of service, and request a service upgrade when the provided service is unsatisfactory. The recovery behavior of the destination-based approach provides slower recovery when compared to the network-based approach introduced by SWAN. A gradual recovery is effective in disrupting a lower number of RT flows, and provides less average variation in RT delays. In addition, ESWAN is shown to decrease the amount of expired RT packets by about 8.5%, which represents a significant improvement compared to the original SWAN implementation.

Most of the mechanisms defined in SWAN and in this research are based on the shaping rate (T seconds). The period T should be small enough to respond to network dynamics, and large enough to average out the high traffic volumes generated by burst traffic [12]. Both SWAN and ESWAN test-beds use the value of T (*2 sec*). Destination nodes can easily assign a value for MAPD based on information from the application layer. The value of MAPD varies considerably, based on the application, but is very essential in customizing the QoS requirements for every flow.

The destination-based approach introduces EBR (β), and EDR (δ) as two important parameters to measures real-time flow quality. The two parameters need to be set up at the session start using probe requests. User satisfaction is a major factor in defining acceptable thresholds for both parameters. For instance, streaming real-time flows may be able to tolerate larger jitter buffers than

interactive real-time flows; as a result, we expect (δ) values to be more stringent for interactive real-time flows. Therefore, further research needs to be done on the ideal and acceptable values of both parameters.

Another important factor in evaluating the SWAN model is investigating the pragmatic approach for bandwidth utilization. SWAN follows a conservative view of bandwidth availability when admitting new real-time flows assuming that the remaining slack of bandwidth may be used by best-effort traffic, or may be used to compensate for deterioration in radio link quality. Therefore, SWAN achieves relatively high resource utilization in installations that has equivalent real-time and best-effort volume of traffic and in installations with limited variations in radio link quality. Additional research is required to evaluate and tune SWAN for environments with skewed percentages of traffic types and highly variable radio link quality. Also a comparison with Shah [14] approach is very important in evaluating the end to end performance.

References

[1] Gahng-Seop Ahn, Andrew T. Campbell, Andras Veres, and Li-Hsiang Sun, *"SWAN: Service Differentiation in Stateless Wireless Ad-hoc Networks"*, Proceeding of the IEEE Infocom, June 2002.

[2] Gahng-Seop Ahn, Andrew T. Campbell, Andras Veres, and Li-Hsiang Sun, *"Supporting Service Differentiation for Real-Time and Best-Effort Traffic in Stateless Wireless Ad Hoc Networks (SWAN)"*, IEEE Transactions on Mobile Computing, September 2002.

[3] Gahng-Seop Ahn, Andrew T. Campbell, Andras Veres, and Li-Hsiang Sun, *"SWAN"*, IETF draft, MANET workgroup, October 2002.

[4] Josh Broch, David B. Johnson, Yih-Chun Hu, and Jorjeta Jetcheva, *"A Performance Comparison of Multi-hop Wireless Ad-hoc Networking Routing Protocols"*, IEEE Proceeding of the ACM/IEEE 4th. International Conference on Mobile Computing and Networking (MobiCom' 98), Dallas TX, USA, October 1998.

[5] Shigang Chen, and Klara Nahrstedt, "*Distributed Quality-of-Service Routing in Ad Hoc Networks*", IEEE Journal on Selected Areas in Communication, vol. 17, no. 8, pp. 1488-1505, August 1999.

[6] D. Chiu and R. Jain, *"Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks"*, Computer Networks, 1989.

[7] S. Corson, and J. Macker, *"Mobile Ad-hoc Networks: Routing Protocol Performance Issues and Evaluation Considerations"*, IETF RFC 2501, January 1999.

[8] David B. Johnson, David A. Maltz, Yih-Chun Hu, and Jorjeta G. Jetcheva, *"The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks (DSR)"*, IETF draft, MANET workgroup, February 2002.

[9] Seoung-Bum Lee, Gahng-Seop Ahn, Xiaowei Zhang and Andrew T. Campbell, *"INSIGNIA: An IP-Based QoS framework for Mobile Ad-hoc Networks"*, Journal of Parallel & Distributed Computing, vol. 60, no. 4, pp 374-406, April 2000.

[10] Mohammad Mirhakkak, Nancy Schult, and Duncan Thomson *"Dynamic QoS and Adoptive Applications for*

variable Bandwidth Environment", MITRE-DoD project paper, URL: <http://www.mitre.org/support/papers/archive99_00.shtml>, April 2000.

[11] Y. Morgan, and T. Kunz, *"PYLON: An Architectural Framework for Ad-hoc QoS Interconnectivity with Access Domains"*, Proceedings of the 36th Hawaii International Conference on System Sciences, Hawaii USA, January 2003.

[12] Charles E. Perkins and Elizabeth M. Royer, *"Ad-Hoc On Demand Distance Vector Routing"*, Proc. IEEE Workshop Mobile Computing Systems and Applications, Feb. 1999.

[13] Charles E. Perkins, Elizabeth M. Belding-Royer, and Samir R. Das, *"Ad hoc On-Demand Distance Vector (AODV) Routing"*, IETF draft, MANET workgroup, June 2002.

[14] Samarth H. Shah, Kai Chen, Klara Nahrstedt, "*Dynamic Bandwidth Management for Single-hop Ad Hoc Wireless Networks*", ACM/Kluwer Mobile Networks and Applications (MONET) Journal, Special Issue on Algorithmic Solutions for Wireless, Mobile, Ad Hoc and Sensor Networks, vol. 10, num. 1, 2005.

[15] Joao L. Sobrinho and A. S. Krishnakumar, *"Quality-of-Service in Ad hoc Carrier Sense Multiple Access Networks"*, IEEE Journal on Selected Areas in Communications, Vol. 17, No. 8, pp. 1353-1368, August 1999.

[16] Andras Veres, Andrew T. Campbell, Michael Barry and Li-Hsiang Sun, *"Supporting Service Differentiation in Wireless Packet Networks Using Distributed Control"*, IEEE Journal of Selected Areas in Communications, Special Issue on Mobility and Resource Management in Next-Generation Wireless Systems, Vol. 19, No. 10, pp. 2094-2104, October 2001.

[17] Hannan Xiao, Winston K.G. Seah, Anthony Lo, and Kee Chaing, "*Flexible QoS Model for Mobile Ad-hoc Networks*", IEEE Vehicular Technology Conference, VTC 2000-spring, IEEE 51st, vol. 1, pp 445-449, Tokyo, May 2000.

[18] Yuan Xue, Baochun Li, Klara Nahrstedt, "*Optimal Resource Allocation in Wireless Ad Hoc Networks: A Price-based Approach*", to appear in the IEEE Transactions on Mobile Computing, 2005.

[19] Network Simulator version 2 <URL: http://www.isi.edu/nsnam/ns/>.

[20] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE Standard 802.11, June 1999.

[21] H. Mantar, J. Hwang, I. Okumus, and S. Chapin "*A Scalable Model for Interbandwidth Broker Resource Reservation and Provisioning*", in the IEEE Journal on Selected Areas in Communications, vol. 22, no. 10, December 2004.

[22] T. Korkmaz and J. Guntaka "*State-Path Decoupled QoS-based Routing Framework*", in the Proceedings of the IEEE Communication Society, Global Communication Conference, Globocom 04, vol. 3, pp. 1515-1519, Dallas, TX, USA, December 2004.

[23] A. Riedl "*A Hybrid Genetic Algorithm for Routing Optimization in IP Networks Utilizing Bandwidth and Delay Metrics*", in the IEEE Workshop on IP Operations and Management, IPOM 02, pp. 166-170, Dallas, TX, USA, October 2002.

[24] Y. Chen, R. Hwang and Y. Lin "*Multipath QoS Routing with Bandwidth Guarantee*", in the Proceedings of the IEEE Communication Society, Global Telecommunications Conference, Globocom 01, vol. 4, pp. 2199-2203, San Antonio, TX, USA, November 2001.

Dr. Y. Morgan received his Bsc. and Msc. from Cairo University and his PhD. from Carleton University in 2005 in the field of Computer Science. He is focused on research related to ad-hoc wireless and mobile networks, sensor networks, and spawning networks. Currently he is part of the IEEE workgroup for 802.11p and P1609.1/.2/.3/ & .4. He is also part of the Dedicated Short Range Consortium (DSRC) and he teaches courses in the School of Information Technology, Carleton University. Dr. Morgan's' authored and co-authored several research papers and contributed to various standard organizations like the IEEE-SA and the IETF.

Dr. T. Kunz received a double honors degree in Computer Science and Business Administration in 1990 and the Dr. Ing. Degree in Computer Science from the Technical University of Darmstadt, Federal Republic of Germany, in 1994. He is currently an Associate Professor in Systems and Computer Engineering at Carleton University, his research interests are primarily in the area of wireless and mobile computing. The main thrust is to facilitate the development of innovative next-generation mobile applications on resource-constraint, hand-held devices, exploring the required network architectures (MANETs, wireless mesh networks, wireless sensor networks), network protocols (routing, Mobile IP, QoS support), and middleware layers. He teaches courses both at the university and to industry, authored or co-authored over 100 technical papers, received a number of awards, and is involved in national and international conferences and workshops.