

Evaluation of the INSIGNIA Signaling System

Seoung-Bum Lee, Gahng-Seop Ahn, Xiaowei Zhang, and Andrew T. Campbell

COMET Group, Center for Telecommunication Research,
Columbia University, New York, NY 10027
insignia@comet.columbia.edu

Abstract. We present an evaluation of the INSIGNIA signaling system that supports quality of service in mobile ad hoc networks. INSIGNIA is based on an in-band signaling and soft-state resource management paradigm that is well suited to supporting mobility with end-to-end quality of service in highly dynamic environments where the network topology, node connectivity and radio channel conditions are time-varying. We evaluate the performance of the signaling system and show the benefit of our scheme under diverse mobility, traffic and channel conditions.

1 Introduction

Mobile ad hoc networks are autonomous distributed systems that comprise a number of mobile nodes connected by wireless links forming arbitrary time-varying wireless network topologies. Mobile nodes function as hosts and routers. As hosts, they represent source and destination nodes in the network while as routers, they represent intermediate nodes between a source and destination providing store-and-forward services to neighboring nodes. Nodes that constitute the wireless network infrastructure are free to move randomly and organize themselves in arbitrary fashions. Therefore the wireless topology that interconnects mobile hosts/routers can change rapidly in unpredictable ways or remain relatively static over long periods of time. These bandwidth constrained multi-hop networks typically support best effort voice and data communications where the achieved “goodput” is often lower than the maximum radio transmission rate after encountering the effects of multiple access, fading, noise, and interference, etc. In addition to being bandwidth constrained mobile ad hoc network are power-constrained because network nodes rely on battery power for energy. Providing suitable quality of service (QOS) support for the delivery of real-time audio, video and data in mobile ad hoc networks presents a number of significant technical challenges.

There has been little research in the area of supporting quality of service in mobile ad hoc networks. What work exists tends to be based on distributed scheduling algorithms [1] that address re-scheduling when the network topology changes and QOS-based medium access controllers. In [2] and [3], multi-hop multi-cluster packet radio network architectures are proposed. Provisioning quality of service is based on “dynamic virtual circuits” derived from network control and signaling found in ATM networks. This approach relies on a “circuit” model that requires explicit connection management and the establishment of “hard-state” in the network prior to communications. We believe there is a need to investigate alternative network models

other than those based on hard-state virtual circuits. New models need to be more responsive to the dynamics found in mobile ad hoc networks. Delivering end-to-end service quality in mobile ad hoc networks is intrinsically linked to the performance of the routing protocol because new routes or alternative routes between source-destination pairs need to be periodically computed during on-going sessions. The IETF Mobile Ad Hoc Networks (MANET) Working Group [4] recently began to standardize inter-network layer technologies (i.e., routing and support protocols). As such, it is presently focused on standardizing network-layer routing protocols suitable for supporting best effort packet delivery in IP-based networks.

In this paper, we present an evaluation of the INSIGNIA signaling system [5], which is a key component of a broader IP-based QOS framework [6] that supports adaptive services in mobile ad hoc networks. The signaling system plays an important role in establishing, adapting, restoring and terminating end-to-end reservations. Based on an in-band signaling approach, INSIGNIA supports fast reservation, restoration and adaptation algorithms. The structure of the paper is as follows. Section 2 presents an overview of the INSIGNIA signaling system. Following this, we present an evaluation of INSIGNIA in Section 3. We evaluate the suitability of the signaling system to support end-to-end service quality under various traffic, mobility and channel conditions. In Section 4, we make some final remarks.

2 Signaling System Overview

The INSIGNIA signaling system is designed to be lightweight in terms of the amount of bandwidth consumed for network control and to be capable of reacting to fast network dynamics such as rapid host mobility, wireless link degradation, intermittent session connectivity and end-to-end quality of service conditions. In what follows, we provide an overview of the signaling system. For full details of the INSIGNIA QOS framework and its signaling specification see [5] and [6], respectively.

2.1 In-Band Commands

Protocol commands support fast reservation, restoration and adaptation mechanisms. These commands are encoded using the IP option field and include *service mode*, *payload type*, *bandwidth indicator* and *bandwidth request* fields. By adopting an IP option in each IP packet header the complexity of supporting packet encapsulation inside the network is avoided.

When a source node wants to establish a fast reservation to a destination node it sets the reservation (RES) mode bit in the IP option service mode field of a data packet and sends the packet toward the destination. On reception of a RES packet, intermediate routing nodes execute admission control to accept or deny the request. When a node accepts a request, resources are committed and subsequent packets are scheduled accordingly. In contrast, if the reservation is denied packets are treated as best effort (BE) mode packets. In the case where a RES packet is received and no resources have been allocated the admission controller attempts to make a new reservation. This condition commonly occurs when flows are re-routed during the lifetime of an on-going session due to host mobility. When the destination receives a

RES packet it sends a QOS report to the source node indicating that an end-to-end reservation has been established.

The service mode indicates the level of service assurances requested in support of adaptive services. The interpretation of the service mode, which indicates a RES or BE packet, is dependent on the payload type and bandwidth indicator, respectively. A packet with the service mode set to RES and bandwidth indicator set to MAX or MIN is attempting to set-up a “max-reserved” or “min-reserved” service, respectively. The bandwidth requirements of a flow are carried in the bandwidth request field. A RES packet may be degraded to BE service in the case where re-routing or insufficient resources exist along the new/existing route. Note that a BE packet requires no resource reservation to be made. The IP option also carries an indication of the payload type, which identifies whether the packet is a base QOS (BQ) or enhanced QOS (EQ) packet. Using the ‘packet-state’ (service mode/payload type/bandwidth indicator) one can determine what component of the flow was degraded. Reception of a BE/EQ/MIN or RES/BQ/MIN packet indicates that the enhanced QOS packets have been degraded to best effort service. By monitoring the packet-state the destination node can issue scaling/drop commands to the source node.

2.2 Fast Reservation

To establish adaptive flows, source nodes initiate reservations by setting the appropriate field in the IP option in data messages before forwarding “reservation request” packets on toward destination nodes. A reservation request packet is characterized as having the service mode set to RES, payload set to BQ/EQ and bandwidth indicator to MAX/MIN and a valid bandwidth requirements. Reservation packets traverse intermediate nodes executing admission control modules, allocating resources and establishing flow-state at all intermediate nodes between source-destination pairs. A source node continues to send reservation packets until the destination node completes the reservation set-up phase by informing the source node of the status of the flow establishment phase using QOS reporting. When a reservation is received at the destination node, the signaling module checks the flow establishment status. The status of the reservation phase is determined by inspecting the IP option field service mode, which should be set to RES. If the bandwidth indication is set to MAX, this implies that all nodes between a source-destination pair have successfully allocated resources to meet the base and enhanced bandwidth needs in support of the max-reserved service mode. On the other hand, if the bandwidth indication is set to MIN this indicates that only the base QOS can be currently supported (i.e., min-reserved mode). In this case, all reservation packets with a payload of EQ received at the destination will have their service level flipped from RES to BE by the bottleneck node.

2.3 Soft-State Management

Reservations made at intermediate routing nodes between source and destination pairs are driven by soft-state management. A soft-state approach is well suited for management of resources in dynamic environments where the path and reservation associated with a flow may change. The transmission of data packets is strongly

coupled to maintenance of flow-states (i.e., reservations). In other words, as the route changes in the network, new reservations will be automatically restored by a restoration mechanism. Once admission control has accepted a request for a new flow, soft-state management starts a soft-state timer associated with the new or re-routed flow. The soft-state timer is continually refreshed as long as packets associated with a flow are periodically received at intermediate routers. In contrast, if packets are not received (e.g., due to re-routing) then the soft-state is not refreshed but times out with the result of deallocating any resources. Since data packets are used to maintain the state at intermediate nodes we couple the data rate of flows to the soft-state timer value. A major benefit of soft-state is that resources allocated during the reservation phase are automatically removed in a fully independent and distributed manner when the path changes.

2.4 Fast Restoration

Flows are often re-routed within the lifetime of on-going sessions due to host mobility. The goal of flow restoration is to re-establish reservation as quickly and efficiently as possible. Rerouting active flows involves the routing protocol [7] [8] [9] (to determine a new route), admission control and resources reservation for nodes that belong to a new path. Restoration procedures also call for the removal of old flow-state at nodes along the old paths. In an ideal case, the restoration of flows can be accomplished within the duration of a few consecutive packets given that an alternative route is cached. We call this type of restoration “immediate restoration”. If no alternative route is cached the performance of the restoration algorithm is coupled to the speed at which the routing protocols can discover a new path. When an adaptive flow is re-routed to a node where resources are unavailable, the flow is degraded to best effort service. Subsequently, downstream nodes receiving these degraded packets do not attempt to allocate resources or refresh the reservation state associated with the flow. In this instance the state associated with a flow is timed out and resources deallocated. A reservation may be restored if resources are free up at a “bottleneck” node or further re-routing allows the restoration to complete. We call this type of restoration “degraded restoration”. A flow may remain degraded for the duration of the session never restoring which we describe as “permanent degradation”. The enhanced QOS component of an adaptive flow may be degraded to best effort service (i.e., min-reserved mode) during the flow restoration process if the nodes along the new path can only support the minimum bandwidth requirement.

2.5 QOS Reporting

QOS reporting is used to inform source nodes of the on-going status of flows. Destination nodes actively monitor on-going flows inspecting status information (e.g., bandwidth indication) and measuring the delivered QOS (viz. packet loss, delay, throughput, etc.). QOS reports are also sent to source nodes to complete the reservation phase and on a periodic basis to manage end-to-end adaptation. QOS reports do not have to travel on the reverse path toward a source. Although QOS reports are generated periodically according to the applications’ sensitivity to the service quality, QOS reports are sent immediately when required (i.e., typically

actions related to adaptation). In the case where only BQ packets can be supported, as is the case with the min-reserved mode service, the signaling systems at the source “flips” the service mode of the EQ packets from RES to BE with all “degraded” packets sent as best effort. Partial reservations [6] that exist between source and destination nodes are automatically timed out after “flipping” the state variable in the EQ packets. Since there is a lack of EQ packets with the RES bit set at intermediate routers any associated resources are released allowing other competing flows to contend for these resources.

2.6 Adaptation

QOS reports are used as part of the on-going adaptation process that responds to resource changes in mobile ad hoc networks. Flow reception quality is monitored at the destination node and based on a user supplied policy, actions are taken to adapt flows under certain observed conditions. Action taken is conditional on what is programmed into the adaptation policy by the user. INSIGNIA provides a set of adaptation levels that can be selected. Typically, an adaptive flow operates with both its base and enhanced components being transported with resource reservation. Scaling flows down depends on the adaptation policy selected. Flows can be scaled down to their base QOS delivering enhanced QOS packets in a best-effort mode hence releasing any partial reservation that may exist. On the other hand, the destination can issue a drop command to the source to drop enhanced QOS packets (i.e., the source stops transmitting enhanced QOS packets). Further levels of scaling can force the base and enhanced QOS packets to be transported in best effort mode. In both cases, the time scale over which the adaptation actions occur is dependent on the application itself. These scaling actions could be instantaneous or based on a low-pass filter operation.

3. Evaluation

3.1 Simulation Environment

The INSIGNIA simulator consists of 19 ad hoc nodes. Each mobile node has a transmission range of 50 meters and shares a 2 Mbps air interface between neighboring mobile nodes within the transmission range. Time-varying wireless connectivity between nodes is modeled using 42 links. The mobility model is based on link failure and recovery characteristics; that is, connectivity is randomly removed and recovered with an arbitrary exponential distribution. Typically, mobile ad hoc networks do not have full connectivity between all mobile nodes at any given time due to the mobility behavior of mobile nodes and time-varying wireless link characteristics. With this in mind, maximum network connectivity is set at 85% such that 15% of the mobile nodes within their transmission ranges remain disconnected.

The architectural [6] components implemented in our simulator include: (i) the signaling system; (ii) the Temporally Ordered Routing Algorithm (TORA) [7]; (ii) a packet scheduler based on a deficit round robin implementation; and (iii) an admission controller, which is simply based on peak rate bandwidth allocation.

For simulation purposes 10 adaptive flows with different bandwidth requirements ranging from 75-500 kbps are operational throughout the simulation. An arbitrary number of best effort flows are randomly generated to introduce different loading conditions distributed randomly throughout the network (i.e., in different parts of the network) during the simulation. We also chose an arbitrary traffic pattern/load with average packet size of 2 Kbytes. Identical traffic/load is used for all scenarios under investigation. The base QOS component of adaptive flows corresponds to 50-70% of an adaptive flow's bandwidth needs whereas enhanced QOS corresponds to 30-50%. For example, an adaptive flow of 300 kbps may have a base and enhanced QOS of 150 kbps for each component such that minimum and maximum bandwidth requirement is set to 150 and 300 kbps, respectively.

The mobility model used throughout the simulation supports three different rates of mobility. Moderate mobility represents slow vehicular mobility ranging from 9-18 km/hr. Mobility conditions slower than moderate mobility is defined as slow mobility (i.e., speed less than 9km/hr) while rates faster than moderate mobility are categorized as fast mobility (i.e., speed exceeding 18km/hr). We inherit the mobility model used in the TORA simulation [10].

For the purpose of evaluating our signaling scheme, we measure per-session and aggregate network conditions for a number of experiments and analyze flow restoration, soft-state management and host/router mobility. We observe throughput, delays, out-of-order sequence packets, lost packets, percentage of delivered degraded packets for the different mobility rates and system configurations (e.g., changing soft-state timers). We are particularly interested in percentage of reserved and degraded packets delivered to receivers. This metric represents the ability of our framework to deliver assurance. We also observe the number of re-routing, degradation, restoration and adaptation events that took place during the course of each experiment as a measure of the dynamics of the system under evaluation.

3.2 Restoration Analysis

In the following experiment we investigate the impact of re-routing and restoration on adaptive flows. Since re-routing of flows requires admission control, resource allocation, state creation and removal of old state, we track the re-routing and restorations events and any degradation that takes place.

Typically, adaptive flows experience continuous re-routing during their session holding times. This is certainly the case with flows that represent continuous audio and video flows but not necessarily the case for micro-flows. Flows may be re-routed over new paths that have insufficient resources to maintain the required QOS. A key challenge for restoration is the speed at which flows can be restored. This is dependent on the speed at which new routes can be computed by the routing protocol if no alternative routes are cached and the speed at which the signaling system can restore reservations. The speed at which old reservations are removed is a direct function of the soft-state timer. The mobility rate impacts the number of restorations observed in the system and therefore the QOS delivered by INSIGNIA. As the rate of mobility increases (e.g., from moderate to fast mobility) restoration algorithms need to be scalable and highly responsive to such dynamics in order to maintain end-to-end QOS.

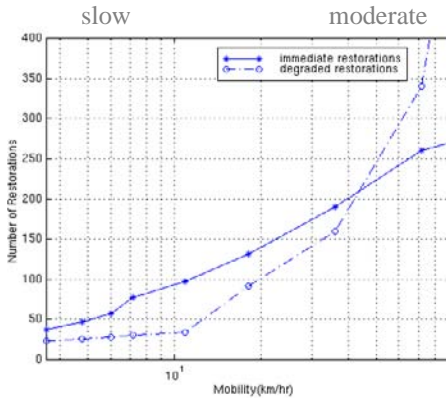


Fig. 1(a). Number of Restorations

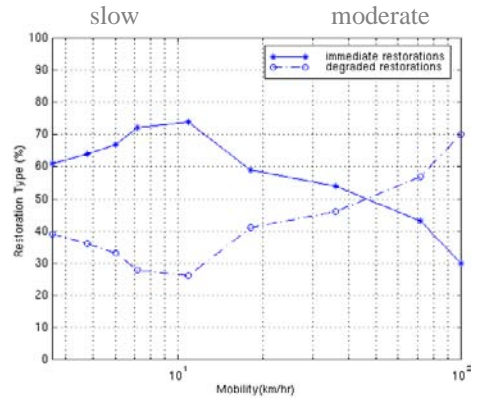


Fig. 1(b). Percentage of Restorations

Figures 1(a) and 1(b) illustrate immediate and degraded restorations observed under various mobility conditions. As indicated in these figures an increase in network dynamics increases the number of “immediate” and “degraded” restorations. The network experiences a total of 38 (61%) immediate restorations and 24 (39%) degraded restorations in the course of the simulation for a mobility rate of 3.6 km/hr as shown in Figure 1(a). As the rate of mobility increases, the ratio between immediate restoration and degraded restoration changes. More immediate restorations are observed in comparison to degraded restorations for slow and moderate mobility conditions as shown in Figure 1(b). However, when mobility conditions exceed 45 km/hr, degraded restoration becomes dominant as shown in Figure 1(b). The connectivity between mobile nodes becomes problematic as the mobility of nodes increases thereby causing the network topology to change rapidly. Consequently, the number of available routes between source and destination nodes diminishes and the contention for network resources increases. This phenomenon introduces service fluctuations and degradation. The INSIGNIA framework adopts a simple admission control test that does not favor re-routed flow over existing flows. A re-routed flow is denied restoration along a new route when insufficient resources are available to meet its minimum bandwidth requirement. This approach minimizes any service disruptions to existing flows preventing a wave of service fluctuation propagating throughout the network. When a mobile host loses its connectivity to neighboring nodes due to mobility, any reservations along the old path are automatically removed. In the case of degraded restoration or permanent degradation, flows are degraded to min-reserved mode or best effort mode because of the lack of resources to restore the flows during re-routing.

We observed that max-reserved adaptive flows are more likely to be degraded to best effort service than are the min-reserved mode adaptive flows. This is mainly due to the admission control policy adopted and semantics of base QOS and enhanced QOS components of flows where the base QOS of a typical adaptive flow consists of 50-70% of the overall bandwidth needs. The admission controller will attempt to support the base and enhanced bandwidth needs of flows. This leads to a situation where most mobile nodes mainly support max-reserved mode flows and a few min-reserved mode flows to fill the remaining unallocated bandwidth. This leads to the blocking of max-reserved flows and due to this behavior the vast majority of degraded flows are “max-reserved to best effort”. Therefore, degraded restorations of

“best effort to min-reserved” (meaning that the min-reserved flow is degraded to best effort before being restored to min-reserved) only occur when the re-routed adaptive flows encounter resources to support min-reserved service. We observed that degraded restoration for “best effort to max-reserved” is the most dominant degraded restoration type observed as shown in Figure 2. This is because re-routed flows are more likely to be accepted or denied rather than degraded to min-reserved flows under slow and moderate mobility conditions. However, we observe that when the mobility is fast that “best effort to min-reserved” degraded restoration becomes the dominant type as shown in Figure 2. In the case of high mobility, only a limited number of routes exist to route flows causing service degradation. The rapid fluctuations in the monitored QOS causes the adaptation process at the destination to request that the degraded flows be scaled down to their min-reserved mode. In this instance, the “best effort to min-reserved” restoration becomes the dominant type as shown in Figure 2.

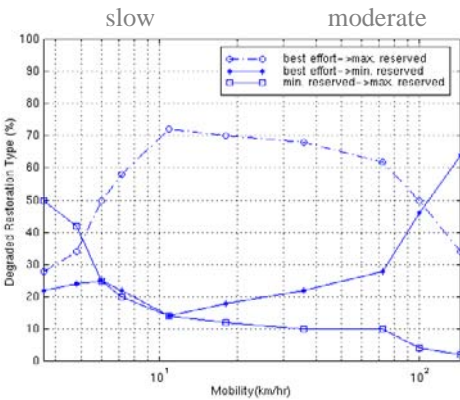


Fig. 2. Degradeestorations Types

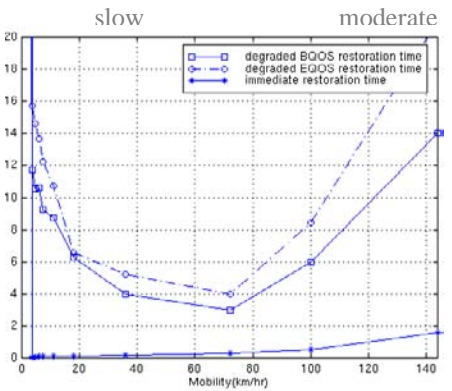


Fig. 3. Time Spent for Immediate Restorations and Degraded Restorations

Increased mobility forces mobile hosts to adapt flows to their min-reserved modes preventing adaptive flows from scaling back up due to the fast time scale dynamics and re-routing experienced. When the mobility is fast, all adaptive flows scale down to their min-reserved service 90 seconds into the simulation trace but only two scale-up adaptations actions are observed during the complete trace. The number of “best effort to max-reserved” and “min-reserved to max-reserved” degraded restorations decreases as mobility is increased to fast as shown in Figure 2. The “best effort to min-reserved” degraded restorations continue to increase implying that most flows scale down to their minimum requirements and operate at the min-reserved mode. Figure 3 illustrates the restoration times across the complete mobility range. The base QOS restoration time corresponds to the time taken to regain the min-reserved service for a flow that has been temporarily degraded to best effort mode. The enhanced QOS restoration time corresponds to the time taken for the max-reserved service to restore from best effort service or from min-reserved service. We observe that the average restoration time required for immediate restoration is relatively constant at 0.2 ~ 0.9 sec under all mobility conditions. In this case immediate restorations only require an interval of two consecutive packets to restore a reservation. However, unlike the

immediate restorations mobility conditions impact the average degraded restoration times as shown in Figure 3.

3.3 Soft-State Analysis

Mobile soft-state management is used to maintain reservations. The duration of soft-state timers has a major impact on the utilization of the network. Figure 4 shows the impact of soft-state times on network performance in terms of the number of reserved mode packets delivered. Reception of a reserved mode packet (with the service mode set to RES) at the destination indicates that the packet is delivered with max-reserved or min-reserved assurance. Reception of a degraded packet implies that the packet has been delivered without such guarantees. Therefore the percentages of reserved and degraded packets received by destination nodes as a whole indicate the degree of service assurance that an INSIGNIA network can support for different values of soft-state timers.

In what follows, we discuss the impact of mobile soft-state timers on network performance. We set the mobile soft-state timer value in the range of 0.01 to 30 sec and observe the corresponding system performance. For each experiment we set the same timer value at each node. As shown in Figure 4, the soft-state timer value has an impact on the overall network performance. The ability to support adaptive services decreases as the soft-state timer value increases. The percentage of delivered reserved packets decreases as mobile soft-state timer increases. The percentage of degraded packets increases as the soft-state timer value increases as shown in Figure 15. Worst case performance is observed when the soft-state timer value is set to 30 sec. In contrast, the best performance is observed when soft-state timer is set to 2 sec as shown in Figure 4. We observed that 69% of the packets are delivered as reserved packets and 31% as best effort packets when the soft-state timer is set to 30 sec. Support for QOS substantially improves with 88% of reserved packets being delivered to receivers with a soft-state timer value of 2 sec. Large timeout values tend to lead to under utilization of the network because resources are “locked up” where resources remain allocated long after flows have been re-routed. New flows are unable to use these dormant resources resulting in the overall degradation of the network due to “resource lock-up”.

As the value of the soft-state timer gets smaller fewer resource lock ups are observed and utilization increases. However, when the timer is set to a value smaller than 2 sec the network experiences what we describe as “false restoration”. This occurs when a reservation is prematurely removed because of a small soft-state timer. However, this is a false state because the session holding time is still active and the source node keeps sending packets. In this case, the reservation is removed because of a timeout and then immediately reinstated when the next reserved packet arrives. False restorations occur when the timeout value is smaller than the inter-arrival time between two consecutive packets associated with a flow. With a soft-state timer of 0.04 sec, for example, all the adaptive flows experienced numerous false restorations. Mobile routers often de-allocate and reallocate resources without the involvement of any network dynamics due to mobility. In the worst case, every packet can experience a false restoration. Such events not only increase the processing costs of state creation and removal, and resource allocation and de-allocation but also falsely reflect the resource utilization and availability of the system. When the network experiences

numerous false restorations, re-routed flows often find nodes with few resources allocated on the new path. This phenomenon causes flows to always gain max-reserved mode resources with mobile nodes accepting the request for resources well beyond their actual capacity. This results in reserved packets experiencing indefinite delays at intermediate nodes even though resource assurances are provided by admission control resulting in wide scale packet losses and service degradation. Figure 4 shows a “false restoration region” where there is little distinction between reserved and best effort operational modes and where reservations are typically always granted. Adaptation and restoration algorithms can fail under false restoration conditions due to perception of unlimited resource availability. Setting a suitable soft-state timer value is therefore essential to preventing both false restoration and resource lock-up in our framework.

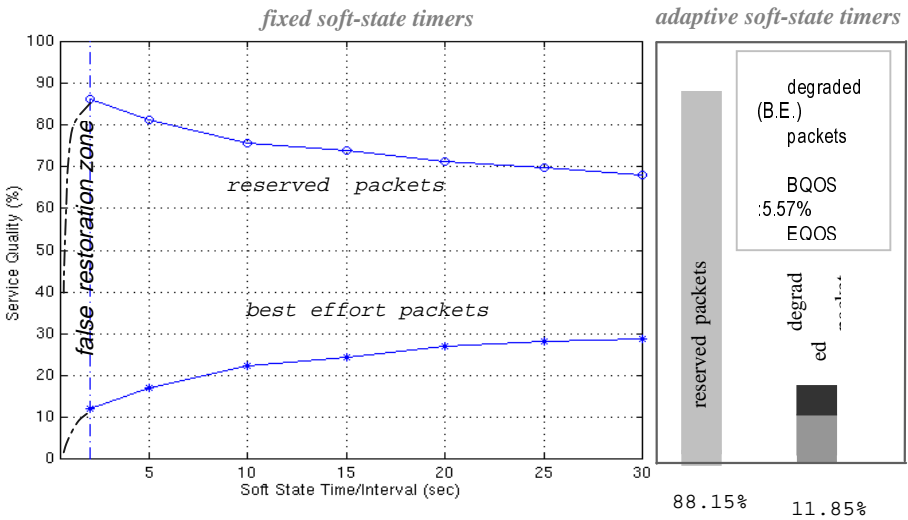


Fig. 4. Soft-State Timers and Network Performance

Each data packet associated with a reserved flow is used to refresh soft-state reservations. We observe that different adaptive flows have different data rates thus a fixed timeout value is too limiting. For example, one value may be fine for some set of flows but cause false restorations or resource lock-up for others. Clearly there needs to be a methodology for determining the value of the soft-state timer. The issue of false restoration and resource lock-up can be only resolved by adjusting the timeout values based on measured flow dynamics. The timeout should be based on the effective data rate of each flow. More specifically, the soft-state timer should be based on the measured packet inter-arrival rate of adaptive flows. The signaling system measures packet inter-arrivals and jitter at each mobile node for each flow, adjusting the soft-state timeout accordingly. In the experimental system we implemented an adaptive soft-state timer that is initially set to 4 sec representing an initial safety factor. This allows mobile nodes to set their soft-state timers according to their effective data rate allowing the timeout to adjust to network dynamics and the variation in the inter-arrival rates of individual flows traversing nodes. The implementation of an adaptive soft-state timeout effectively removes resource lock-

ups and false-restorations as shown in Figure 4. We observe that when an adaptive soft-state timer scheme is adopted 86% of flows are delivered as reserved packets, 11% as degraded packets and 3% of packets are lost. Adaptive soft-state timers greatly reduce resource lock up and false restoration conditions allowing the network to support better service assurances through the delivery of more reserved packets and fewer degraded packets at destination nodes.

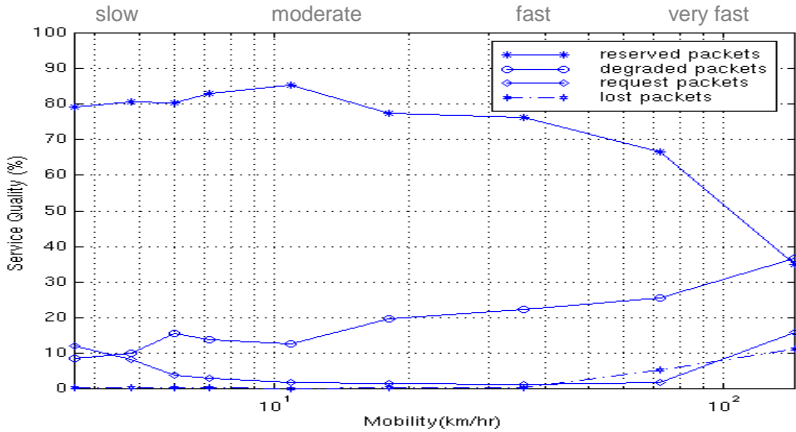


Fig. 5. Mobility and Network Performance

3.4 Mobility Analysis

To evaluate the impact of mobility on the INSIGNIA, we conduct a set of experiments operating under identical traffic pattern/load conditions and various mobility conditions ranging from 0 km/hr to 72 km/hr. Figure 5 shows the impact of mobility on the delivered service quality. When there is no host mobility, results closely approximates a fixed network infrastructure where admitted flows receive stable QOS assurances. One anomaly is observed, however. Six adaptive flows failed to be granted reservations due to lack of network resources at intermediate nodes. As a consequence only 49% of the packets are delivered as reserved packets and 51% as best effort packets. This anomaly is a product of the routing protocol, which provides a non-QOS routing solution. Adaptive flows are routed to bottleneck nodes resulting in the failure of admission control due to the lack of resources. This problem could be resolved by designing a signaling system that takes alternative routes in the case that admission control fails along a selected path.

With the introduction of mobility into the network, the performance improves (i.e., more reserved packets are delivered) as illustrated in Figure 5. Mobility induced re-routing allows request packets to traverse alternative paths increasing the probability of finding a route with sufficient resource availability to admitted flows as reserved mode packets. Figure 5 shows that the INSIGNIA supports relatively constant QOS under slow and moderate mobility conditions between 3.6-18 km/hr. The optimal performance is observed when the average network mobility is approximately 11 km/hr. This results in the delivery of 86% of reserved packets. The

in-band nature of INSIGNIA allows the system to cope with fast network dynamics in a responsive manner. In an ideal case, INSIGNIA requires only a single packet reception to set up and restore (i.e., immediate restorations) reservation for the new or re-routed flows, respectively. INSIGNIA supports the delivery of 66% reserved packets even when mobile hosts are moving at 72 km/hr as shown in Figure 5. This is a very encouraging result.

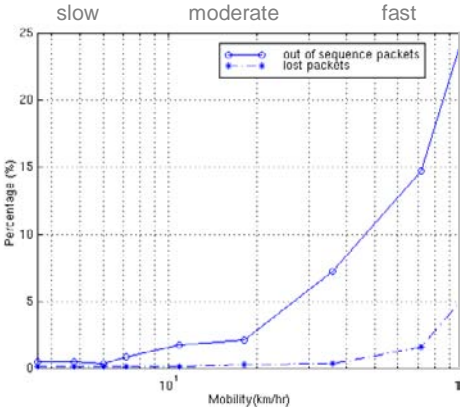


Fig. 6. Impact of Mobility on Out of Order Delivery and Packet Loss

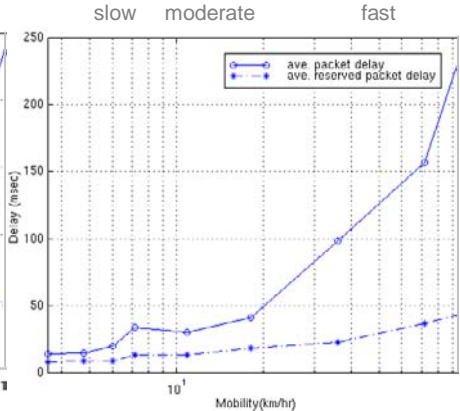


Fig. 7. Packet Delays

Note that the service provided in a mobile ad hoc network has a memoryless property such that adaptive flows require new admission tests along the new path when re-routing occurs. This implies that an increase in mobility may cause fluctuations in perceived service quality. At 72 km/hr all flows are scaled-down to min-reserved packets (after 90 sec into the simulation) due to fluctuations in the delivered quality. At this speed only two flows are capable of regaining their max-reserved service. When mobility conditions exceed 72 km/hr support for QOS breaks down rapidly as indicated in Figure 5. The mobility characteristics overload the system and service assurance for adaptive flows diminish. In fact, when mobility exceeds 90 km/hr, we observe that a number of flows are transported as best effort packets for more than 70 sec because they failed to accomplish their end-to-end flow set up due to persistent loss of RES packets and QOS reports. This phenomenon corresponds to the abrupt loss of reserved packets and degraded packets.

An increase in out-of-sequence packet is also observed at higher speeds possibly causing service disruption at the receiver. Figure 6 shows the number of out-of-sequence packets under various mobility conditions. The number of out-of-sequence packets generally increases as mobility increases. The different propagation delay characteristics of reserved and best effort packets associated with the same end-to-end flow has an impact on out-of-sequence packets. Figure 6 also shows the number of lost packets observed under different mobility conditions. Packets that are delayed for more than 15 sec are discarded at intermediate nodes and are considered to be lost packets. Figure 7 shows the delay characteristics of packets under various mobility conditions. When mobility increases, the connectivity between nodes becomes problematic. Such network dynamics trigger frequent routing updates and decreased

connectivity. Thus, the number of available routes between nodes decreases as mobility increases. Degraded packets queue up at intermediate nodes experiencing long delays. However, the reserved packets are less sensitive to these delays as indicated in Figure 7 with all reserved packets being delivered within a period of 40 msec.

4 Conclusion

In this paper, we have presented the evaluation of the INSIGNIA signaling system, an in-band signaling system that supports fast reservation, restoration and adaptation algorithms that are specifically designed to deliver high performance adaptive services in mobile ad hoc networks. The signaling system is designed to be lightweight and highly responsive to changes in network topology, node connectivity and end-to-end quality of service conditions. Our simulation results show the benefit of INSIGNIA under diverse mobility, traffic and channel conditions in support of fast reservation, restoration and adaptation. The use of in-band signaling and soft-state resource management proved to be very efficient, robust and scalable under a wide variety of network conditions. Our results highlighted a number of anomalies that emerged during the evaluation phase. However, the use of adaptive soft-state timers resolved these issues.

Acknowledgements

This work is supported in part by the Army Research Office under award DAAD19-99-1-0287. We are particularly indebted to Vincent Park and Dr. M. Scott Corson for providing the INSIGNIA project with the source code for TORA.

References

1. Ephremides, A. and T. Truong, Scheduling Algorithm for Multi-hop Radio Networks, *IEEE Trans. on Computers*, 38:1353, 1989.
2. C. R. Lin and M. Gerla, Asynchronous Multimedia Multihop Wireless Networks, in *Proc. IEEE INFOCOM'97*, April 1997.
3. R. Ramanathan and M. Streenstrup, Hierarchically-Organized, Multi-hop Mobile Wireless networks for Quality-of-service Support, Technical Report, BBN, 1998.
4. J. Macker, and M. S. Corson, Mobile Ad hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations, draft-ietf-manet-issues-01.txt, *Internet Draft*, work in progress, April 1998.
5. Lee, S-B and A. T. Campbell, INSIGNIA, *Internet Draft*, draft-lee-insignia-00.txt, work in progress, November 1998.
6. Lee, S.B., Gahng-Seop, A., Zhang, X., 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), April 2000.

7. V. Park, and S. Corson, Temporally Ordered Routing Algorithm (TORA) Version 1 Functional Specification, draft-ietf-manet-tora-spec-00.txt, *Internet Draft*, work in progress, November 1997.
8. C. Perkins, Ad hoc On demand Distance Vector (AODV) Routing, draft-ietf-manet-aodv-01.txt, *Internet Draft*, work in progress.
9. D. B. Johnson and D. A. Maltz, Dynamic Source Routing in Ad Hoc Wireless Network, *Mobile Computing*, Chapter 5, pp. 153-181.
10. TORA OPNET source code supplied by V. Park and M. S. Corson, 1997.