UNIVERSITY OF CALIFORNIA

Los Angeles

Mobility Issues in Hybrid Ad-Hoc Wireless Sensor Networks

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Electrical Engineering

by

Vishal Ailawadhi

2002

The dissertation of Vishal Ailawadhi is approved.

Mario Gerla

Izhak Rubin

Kung Yao

Gregory J. Pottie, Committee Chair

University of California, Los Angeles

2002

DEDICATION

To my family, near and far…

With Love…

Contents

List of Figures

List of Tables

ACKNOWLEDGEMENTS

Inspiration is the root of all accomplishment. It comes in all shapes and sizes, in all places, at all times, and most importantly, when you least expect it. For this reason, I don't have the capacity to begin to express all the people and events which have guided me to this point. I would like to offer, though, a few important thoughts.

First and foremost, none of this would be possible without the support of my parents, Nisha and Manohar Ailawadhi, and my sister and her husband, Ritu and Rajesh Grover. Their encouragement has convinced me that I can achieve anything I put my mind to. In a large part, this work belongs to them as much as it belongs to me.

Here, at UCLA, I had the good fortune of working with Greg Pottie. I couldn't have imagined working for anyone else. Apart from his support as a professor and a researcher, I was fortunate to have his support as a friend. From the engineering perspective, he gave me freedom in developing ideas, and the guidance to reach my goals.

For many of my initial years of research, I was fortunate enough to work with Kathy Sohrabi and Jay Gao on the self-organization network protocols. As a fledgling in any area, it is important to have colleagues to turn to for guidance. From questions

regarding any subject, they were there to simplify all the transitions associated with beginning research.

Finally, I would like to thank all of my friends who have come and gone. Ten years as a student in one university will introduce you to many new and exciting people, far too many to list here. But, it is the everyday interaction with all of these people that got me through this last decade. Many thanks to all of them.

VITA

July 17, 1974	Born, California, USA
1996	B. S. Electrical Engineering Summa Cum Laude University of California, Los Angeles Los Angeles, CA
1997	M. S. Electrical Engineering University of California, Los Angeles Los Angeles, CA

PUBLICATIONS AND PRESENTATIONS

K. Sohrabi, J. Gao, V. Ailawadhi and G. Pottie, "A Self Organizing Wireless Sensor Network," *Proc. 39th Annual Allerton Conf. Communications, Control, and Comp.*, Urbana, IL, Oct. 1999.

K. Sohrabi, J. Gao, V. Ailawadhi and G. Pottie, "Protocols for Self-Organization of a Wireless Sensor Network," *IEEE Personal Communications*, vol. 7, no. 5, Oct. 2000, p. 16-27.

L. Yip, J. C. Chen, V. Ailawadhi, R. E. Hudson, K. Yao, and G. J. Pottie, "Graphical Integrated Multi-Sensor Simulator for a Wireless Sensor Network," *Proc. ASC, Advanced Sensor Consortium*, March 2001, p. 31-35.

ABSTRACT OF THE DISSERTATION

Mobility Issues in Hybrid Ad-Hoc Wireless Sensor Networks

by

Vishal Ailawadhi

Doctor of Philosophy in Electrical Engineering

University of California, Los Angeles, 2002

Professor Gregory J. Pottie, Chair

A Hybrid Ad-Hoc Sensor Network (HANET) is presented, consisting of both mobile and stationary nodes. Wireless sensors in this domain consist of densely distributed, low-power, energy constrained stationary sensors, which are able to form MAC level connections and Network level multi-hop routes at runtime. The limited energy supplies and the interaction of sparse mobile nodes impose stringent requirements for low-complexity, low-energy, distributed protocol design.

The EAR protocol (Eavesdrop and Register) assumes a mobile-centric view of connection maintenance, as the mobile nodes are assumed to have fewer constraints on energy supplies as compared to the stationary nodes. The mobile node forms a registry based on "Eavesdropping" the signals native to the stationary network's MAC protocols.

The density of the stationary network is exploited to avoid handoffs and acknowledgement messages.

We present the MIR (Mobile Intermediate Routing) protocol to handle routing issues for packets associated with mobile nodes. The stationary wireless backbone will facilitate routing, using intelligent packet forwarding and localized route updating to allow packets to be redirected en route to their destinations. Intermediate Rerouting is shown to reduce the packet delay time and the packet dropping rate.

An algorithm is presented for radio control via the MAC layer. We introduce the concept of Internal Message Updating and External Message Updating, the combination of which allows the mobile nodes to reconfirm present connections while searching for new connections, incurring the energy costs associated with radio level signaling. We present an idea to allow mobile sensors to forgo ICU and ECU functionality by using outage prediction and taking advantage of node proximity. It is shows that this scheme can reduce energy consumption while maintaining a high quality of service.

D-PEC (Distributed Pre-Event Clustering) allows the stationary sensor network to form clusters and combine data to track and identify reoccurring targets, such as mobile nodes. This algorithm allows the network to cluster itself prior to target inclusion, thereby avoiding the costs and delays associated with on-the-fly cluster generation. A bound on the clustering signaling complexity per node is derived for the high density and low density network cases.

Chapter 1

Introduction

The introduction of low-power, low cost wireless devices has allowed networks to support functionality beyond simple communications. Traditional wireless networks, composed of primarily immobile units will be replaced with untethered networks supporting a combination of stationary and mobile nodes. It has become possible to envision a new paradigm by which a large group of wireless nodes can participate in tasks such as high-rate mobile multimedia data transfer, environmental sensing and sampling, homeland security and defense, and health monitoring.

The move away from wired communications, such as simple LANs and land-line telephony has sparked interest in a variety of networks, ranging from communications-centric cellular networks [24, 55, 57] and mobile ad-hoc networks (MANETs) [3, 15, 22, 32, 38, 39], to the data-centric wireless LANs [9, 46] and wireless sensor networks [42, 43, 44, 51, 52]. These networks are characterized by benefits such as distributed functionality and the absence of a tethered backbone (except for the stationary base station backbone in cellular networks), and drawbacks, such as limited energy reserves and bandwidth. As a consequence, new protocols must be designed to take advantage of these new network configurations while preserving key resources. Most importantly, as the wired infrastructure is eliminated, these networks are free to support some degree of mobility in a subset, or possibly all, of the participating nodes.

1.1 Ad Hoc Networks Defined

The departure from traditional wired networks has brought about a shift in the primary concerns which need to be addressed when designing protocols of operation for wireless networks. While, like their wired counterparts, wireless networks are interested in high throughput and low-cost design, network priorities will dictate a tradeoff for an increase in power efficiency and bandwidth optimization. In particular, we consider ad-hoc networks. These peer-to-peer networks are composed of tens to hundreds of possibly homogeneous nodes with sensing and processing capabilities. These nodes will communicate through the wireless medium, achieving ranges of communication of up to hundreds of meters. The ad hoc nature of these networks arises as the nodes are either (a) required to configure a MAC level link architecture without the aid of a centralized protocol (stationary network), or (b) required to adjust MAC level connectivity in the face of a varying topology (mobile network). Generally, ad-hoc networks have been analyzed by assuming that all nodes within the network have similar mobile tendencies (either stationary or mobile) [3, 9, 15, 52].

1.2 Hybrid Ad-Hoc Wireless Sensor Networks

We define wireless ad hoc network which provide support for both mobility and stationary sensor nodes as Hybrid Ad-Hoc Wireless Sensor Networks (HANETs). The configuration of the HANETs considered here consists of a densely populated stationary wireless sensor network, with its own abilities and protocols, which is able to support a low degree of mobility within a small subset of nodes. A stationary ad-hoc network is able to form link-level connections, as well as provide multi-hop routing paths to a sink node, without the use of a centralized processing node. On a functional level, these stationary wireless nodes can act as a backbone sensing network, while the mobile nodes can engage in tasks such as widespread environmental sampling or personnel security patrolling.

Homogeneous sensors are usually deployed, with a random dispersion, in a sensor field where wired connections are either not possible, due to inhospitable terrain, or not desired, due to the costs associated with setting up temporary connections. These networks may be used in situations where physical placement and interaction is not possible, such as battlegrounds and remote planetary surfaces. In such cases, it may not be possible to replace energy reserves on these sensor nodes, suggesting that, in the absence of a wired infrastructure, network lifetime will be limited. Thus, to prolong the lifetime of the sensor networks, energy is considered to be a prime resource at the stationary sensor nodes. Protocols have been designed [44, 52] which suggest that the MAC level link formation and the network level multihop routing formation can be achieved in an energy efficient manner.

As mobile nodes are introduced into the stationary network, though, maintaining connectivity to the network, as well as setting up temporary route formation, will increase the drain of energy reserves at the stationary nodes. Their inclusion in HANETs is desired, though, as they extend network functionality. Mobile sensors, for example, can broaden the spatial sampling area of the network, acting as an information source when data has been collected. Military personnel, on the other hand, can be used to extract data, either locally or globally, from any point in the network, acting as an information sink. Furthermore, two mobile nodes may wish to communicate via the wireless stationary backbone created by the stationary portion of the HANET. The goal of mobility support, therefore, is to design protocols which allow for the maintenance of link level network connectivity, as well as packet routing capabilities, while adhering to the energy constraints of the stationary network.

1.3 Summary of Contributions

The work which will be presented here is concerned with the inclusion of mobile sensor nodes in a stationary ad hoc wireless sensor network, generating a HANET. We assume that the stationary network has reached its steady state operation. That is, the stationary nodes have been deployed (possible in a random dispersion), a link level architecture has formed, routing paths from any sensor node to a sink node have been established, and each sensor node is periodically searching for new neighbors to incorporate into the network. Furthermore, the stationary sensors are assumed to be highly energy constrained.

Prior to this work, few models of this network architecture had been reported in the literature. Thus, we proceed to develop a suite of protocols to support mobile node interaction within this energy constrained stationary sensor network, including MAC and

Network level design. Following this, an energy efficient radio control protocol is presented. In cases where the stationary network is to track/identify the mobility within the HANET, as opposed to offer connectivity, a pre-event clustering scheme is presented to reduce signaling overhead when identifying mobile targets.

1.3.1 MAC Layer Protocol Design

The EAR protocol allows the mobile nodes to achieve connectivity to the stationary sensor network, while taking into account the network architecture and the energy constraints at the stationary nodes. The question posed is whether stationary nodes should be allowed to maintain control of handoffs within the network, depleting energy supplies.

At the mobile node, the EAR protocol generates a registry of stationary nodes to which connections may be desired, based on received signal power. We define three new messages to accomplish mobile handshaking. Slot assignments are discussed, along with the ability to forgo handoffs due to the high density of the stationary network. Furthermore, we show that the high network density also provides the ability to avoid acknowledgement messages, instead using timeouts to guarantee message reception. The details of the EAR protocol and accompanying results are presented in Chapter 3.

1.3.2 Network Layer Protocol Design

The MIR protocol offers routing capabilities to mobile nodes which are connected to the stationary network. We assume that the stationary network itself has formed multihop routing trees which lead from any node to a sink, or user, node. We discuss the ability of utilizing the stationary wireless backbone offered by the HANET architecture. Again, protocols are developed to conserve energy at the stationary nodes.

Three type of routing are discussed as motivation for the MIR protocol. This protocol will allow the stationary network to forward packets from mobile nodes to the sink node via its own multihop tree, as well as set up reverse paths for downlink transmission capabilities. Local routing adaptation is possible via the selective forwarding of a control message. Coupling the ideas of packet forwarding and local adaptation, we present a scheme which is able to intermediately reroute packets when the destination mobile node has moved, or in the case of mobile to mobile transmission. The details of the MIR protocol and accompanying results are presented in Chapter 4.

1.3.3 Power Efficient Radio Control

We introduce the concept of ICU (Internal Connection Updating) and ECU (External Connection Updating) as functions which allow the mobile nodes to maintain and form new connections, respectively, at the cost of energy associated with signaling. We describe a radio level interaction scheme between mobile and stationary nodes which takes advantage of node proximity to allow mobile nodes to forgo ICU and ECU functionality by using outage prediction. By incorporating this prediction scheme, the sensor nodes can preserve energy reserves while still maintaining a high quality of service. The details concerning the radio control algorithm are presented in Chapter 5.

1.3.4 Energy Efficient Distributed Pre-Event Clustering

The D-PEC clustering algorithm allows the stationary sensor network to form clusters of nodes, allowing data sharing capabilities, for the purposes of event tracking and identification. As various events, including mobile targets, tend to be repetitive in nature, and possibly spatially correlated in time, it is possible that event based clustering schemes incur a high signaling overhead associated with repeated cluster set-up. Thus, we present a distributed algorithm which allows the network to cluster itself immediately following a MAC level link formation phase.

The D-PEC algorithm follows a three phase operation, in which nodes determine possible cluster candidates, offer invitations, and declare cluster membership. The final characteristics of the clusters are discussed, along with comparisons to other clustering schemes. The bounds on the maximum signaling required by a network to form clusters using D-PEC are analytically derived for the high density and low density network cases. The details of the D-PEC algorithm, along with performance results, are given in Chapter 6. Concluding remarks will be given in Chapter 7.

Chapter 2

Background Material

2.1 Wireless Networks

Wireless networks are configurations of devices in which members (or a subset of members) communicate to each other via the wireless domain. We focus here on radio signal propagation. The classification of these networks depends on various factors, including the application level goals, the number of communicating nodes, the level of node mobility, and the resources available to each node, just to name a few. Here we present a few of the dominant wireless network configurations, ranging for the data-centric WLAN and Sensor Networks to the communications-centric Cellular and MANET.

2.1.1 Wireless Local Access Networks

The growing field of WLANs accompanies a marked detachment from tethered network architectures, such as wired desktop devices. Smaller, possibly mobile, devices such as laptops or PDAs can be used to exchange information with other devices either directly or via a wireless hub in areas where the physical layout of wires is difficult or expensive. Current applications involve inventory control, hospitals and schools (moderately mobile environments), factory and warehouse settings, and historical buildings. Various protocols are being developed in this field, of which the IEEE 802.11 and HIPERLAN standards are most prominent.

The two topologies supported by the 802.11 standard are (a) backbone communications and (b) ad hoc communications. In scenarios involving a backbone, multiple Access Points are deployed strategically ensuring overlap of a coverage area. Mobile stations (members of the Basic Service Set) will communicate to other stations via these Access Points. For ad hoc direct communications, there is no infrastructure, and nodes are able to communicate directly to other nodes, with no multihop routing support. The nodes themselves communicate at low power levels (100 mW or less), but the moderate bit rates (1-2 Mb/s) allows for communication distances of up to 100 meters, depending on the environmental geometry. The physical layer of the 802.11 standard incorporates either frequency-hopped or direct-sequence spread spectrum communications in the 2.4 GHz frequency band. The MAC level architecture supports CSMA with collision avoidance, incorporating a contention window of backoff times when collisions occur. As collision detection is not possible, communicating nodes solve the hidden node problem by broadcasting RTS (request to send) and CTS (clear to send) packets.

The HIPERLAN standard is an attempt by the European community to create a network which would be comparable to Ethernet services. Again, the wireless devices are generally low power (for safety reasons), but the higher data rates (up to 23.529 Mb/s) decrease the communications range to about 10-100 meters. HIPERLAN networks support features such as multihop routing, time bounded services, and power saving techniques. As in the 802.11 standard, a mechanism is in place for the MAC layer to resolve medium contention via a three phase carrier sensing mechanism. Physical layer techniques incorporate error correction coding, bit interleaving, training sequences, and equalization to mitigate the effects of intersymbol interference due to the high bit rate [26, 35].

2.1.2 Cellular Networks

Similar to the backbone architecture of the 802.11 standard, cellular networks allow the communications between mobile devices via access points, called Base Stations. The set of Base Stations, each with ranges on the order of kilometers, forms a fixed, wired infrastructure, facilitating call routing. Base Stations will provide service to thousands of mobile subscribers, each of which is one hop away from any of the fixed points. Features such as “handoff” and “roaming” allow seamless coverage to mobile users as they move from cell to cell, or from area to area. As power consumption is not a concern, especially at the wired Base Stations, the primary goals of cellular network are low signal outage and high bandwidth efficiency.

Developed in the 1970’s at Bell Labs, the Advanced Mobile Phone Service (AMPS) became the first generation cellular standard. AMPS operated in an analog environment, utilizing frequency modulation (FM) for speech transmission, frequency shift keying (FSK) for signaling, and frequency division multiple access (FDMA) for supporting individual calls within the same bandwidth [35]. Three elements were envisioned in early cellular design: wireless terminals, base stations, and switches. The base stations were to exchange radio signals with the wireless terminals, while the switches controlled the assignment of radio channels and provided a connection between the wireless information networks and fixed networks [16].

In the early 1980's, the GSM system (Global System for Mobile communications, or Groupe Spécial Mobile) was developed. As a digital cellular system, it allowed the use of TDMA and CDMA as multiple access techniques, providing support for more users than traditional analog methods. In TDMA, users are designated a frequency (if FDMA is also employed) and a time slot within a communications frame. In CDMA, the entire frequency is allocated to each user, with spreading codes providing orthogonality [35]. In addition to allowing more sophisticated multiple access techniques, digital cellular systems standardized the signaling involved in roaming and handoffs. Also, the responsibility for the handoff functionality was shifted from primarily switch controlled to a combination of mobile, base station, and switch controlled [16].

2.1.3 Mobile Ad Hoc Networks

Mobile ad hoc networks, or MANETs, show a distinct departure from cellular networks and WLANs, in that there is no need for a central access point or base stations. Stretching back to the 1970's as Mobile Packet Radio Networks, MANETs have received considerable attention over the past 25 years. These networks enable an autonomous system of fully mobile nodes to engage in peer to peer communications at distances of up to hundreds of meters. The ad hoc nature of this network arises in that the topology of the network is continuously changing, and the nodes must continuously update MAC level connection architecture and Network level routing tables. Thus, the primary goal for MANETs is to maintain routing abilities and network organization [8, 21].

Each node in a MANET consists of routers and communications devices, with the entire set of nodes forming a mobile routing infrastructure, where the routing backbone is allowed to be mobile, along with the end devices. In some MANET architectures, a separate backbone does not exist, with routing occurring in multihop fashion, whereas other configurations utilize clusterhead nodes to form a hierarchical routing strategy. In either case, a wireless connectivity needs to exist between all the nodes at any point in time.

Certain qualities characterize MANETs over all other networks. Most importantly, due to node mobility and variations in transmission and reception powers, MANETs experience a high level of topological variability. Also, as wireless links have a lower capacity than wired links, MANETs experience constrained bandwidth and delay characteristics. The nodes themselves are assumed to be battery operated, suggesting an energy constrained operation. Finally, the wireless channel itself is more prone to environmental vulnerabilities [8, 21]. Further details of protocols used in MANET and cellular network design will be given in chapters 3, 4, and 5.

2.1.4 Wireless Sensor Networks

The capabilities of wireless sensor networks arise with the convergence of wireless communications, digital electronics, and micro-electro-mechanical systems, enabling the incorporation of sensing, signal processing, and communications in one packaged device. These low-cost, low-powered devices are able to participate in untethered communications at short distances, allowing the collaborative effort of a large number of nodes. The sensor nodes themselves can be deployed randomly over a large area (possibly inhospitable terrain, disaster relief) close to the observable phenomena. Due to the large number of nodes, and the possible random deployment, centralized protocols become infeasible, giving way to distributed protocol design for self-organizing capabilities. Furthermore, nodes have computational resources available to them, suggesting a balance between local, on-board processing and cluster-based data sharing techniques. Applications for wireless sensor networks include health monitoring systems, environmental sampling and surveying, and military and security environments [1, 13, 42, 43].

Wireless sensor networks are closely related to MANETs, but some substantial differences preclude the ability to apply MANET operational techniques here. Most importantly, the nodes in wireless sensor networks are not assumed to be fully mobile, enabling all or a subset of nodes to provide stationary sensing abilities. Also, wireless sensor networks can be composed of multiple thousands of nodes (several orders of magnitude higher than MANETs) which are densely deployed, offering short multihop routes throughout the network. The nodes themselves are considered to be highly energy constrained, as the battery reserves are not easily replenished. Thus, the topology of sensor networks is variable due to both mobility on a subset of nodes and node failures. The primary goal of sensor networks is to prolong the network lifetime, in the face of sensing operations, node failures, and node mobility. Protocols must be designed which enable power conservation at the expense of degradation in throughput and delay characteristics [1].

2.2 Protocol Design

The individual device operation for all nodes within a communications network is divided into layers, with each layer governed by its own rules or protocol. Each protocol layer is standardized to all other devices in the network, allowing layer to layer communication between all nodes. Within the device's protocol stack (hierarchical grouping of layers), one layer can communicate only with the layers immediately above or below itself. Importantly, the layers are not concerned with the other protocols running in the stack, only its own, in addition to the information required to be shared in the adjoining layers. This allows the substitution of protocols at various layers without affecting the others. The most common protocol stack found in wireless communications is composed of (lowest to highest) the Physical layer, the Data Link layer, the Network layer, the Transport layer, and the Application layer. Here, as well as in the next few chapters, we will focus on the MAC (Medium Access Control) layer, which is a subset of the data link layer, and the Network layer [31].

2.2.1 MAC Layer

The MAC sublayer addresses the problem of controlling access of the wireless devices to the transmission medium in a broadcast network. In situations where several devices communicate by sharing the same medium, the MAC protocol will schedule and designate access channels to avoid simultaneous transmissions (collisions) on the same channel. MAC protocols can be divided into two groups: static allocation protocols and dynamic allocation protocols. Here, a brief description of each is given, along with examples of MAC level solutions.

Static allocation protocols attempt to regulate the channel by dividing the available bandwidth into sections for access by a given number of users. One method of doing this is to provide N users with an equal share of the bandwidth, which has been separated into N sections. This may result in an inefficient use of the available bandwidth, though, unless all users intend to transmit an equal share of data at all times. One method of static allocation is Frequency Division Multiplexing (FDM). Here, the available frequency band is subdivided into N subbands, one for each user. Users then have access to their own particular assigned subband at all times. Time Division Multiplexing (TDM), on the other hand, attempts to divide the communications frame into N time slots. Users then have access to the entire bandwidth during their time slot, and must remain silent at all other times. The downside of TDM is that all users must be time synchronized, other wise collisions may occur. In a more general sense, using static allocation, the frame can be divided into N slots, with the frequency band within each slot divided into M subbands, yielding a total number of $M \cdot N$ distinct orthogonal channels [31].

To increase bandwidth utilization efficiency, channels can be assigned on an as-needed or on-demand basis, as in dynamic allocation protocols. ALOHA was one of the earliest dynamic protocols. In this scheme, all users are allowed to transmit whenever data becomes available. As distinct channels are not assigned to each device, collisions will occur frequently. When a collision is detected, a wireless device will wait for a random backoff period before attempting communications again. Pure, or unslotted, ALOHA has a bandwidth efficiency of 18%, whereas slotted ALOHA (transmissions must begin at the beginning of a time slot) yields a bandwidth efficiency of 37%.

To increase bandwidth efficiency, protocols must be able to allow the devices to sense the channel for current transmissions before initiating their own transmissions, as with CSMA (Carrier Sense Multiple Access) techniques. Here, devices will listen to the medium to determine if the channel is available. If the channel is busy, the devices will follow one of three schemes to determine a transmission time. In 1-persistent CSMA, the device will continually monitor the channel until a previous transmission ends before beginning its own (55% efficiency). In non-persistent CSMA, the device will wait a random time before checking the channel again (90% efficiency). In p-persistent CSMA, the device will continuously monitor the channel, as in the 1-persistent case, but will transmit only with probability p. Thus, as p approaches 1, this case approaches the 1-persistent case. But, as p is reduced, a high bandwidth efficiency is experienced at the cost of high delay ($p = 0.01$ implies almost a 100% efficiency). Various other flavors of CSMA exist, with options for collision avoidance (CSMA/CA) and collision detection (CSMA/CD) [31].

Another broad class of wireless MAC level protocols follows the Multiple Access with Collision Avoidance idea (MACA). In many cases, collision avoidance in broadcast applications is accomplished by simply not transmitting data when another node has begun transmission. The classic problem associated with collision avoidance is the Hidden Terminal Problem. In this case, devices **B** and **C** are within the connection range of device **A**, but not in connection range with each other. Node **B**, therefore, will not be able to determine when node **C** has begun transmission to node **A**. Thus, in the MACA scheme, a device wishing to communicate will initially send a RTS (Request to Send) packet, to which the receiving node will reply with a CTS (Clear to Send) packet. This will assure that all nodes within the communications range of both of these nodes will avoid transmissions during this time [31].

2.2.2 Network Layer

The Network layer ensures that all packets are properly routed from the transmitter to the receiver. In point to point networks, like ad hoc networks, this problem become complex, as the protocol must be able to choose routes for packets based on such factors as (a) net hop count from sender to receiver, (b) overall distance traveled by the packet, (c) local network congestion avoidance, and (d) available resources at the nodes on the potential route [31]. Presented here is a brief description of the techniques used in route formation in networking protocols.

Routing algorithms can be classified based on their method of routing table maintenance, either Table-Driven or On-Demand. Table Driven protocols are similar to the connectionless approach of packet forwarding. All nodes in the network are made aware of all of the routes leading to all other nodes with the network. Thus, during the lifetime of the network, signaling must occur to continuously update the routing tree information at each node (with large packets to redefine routes, and small packets to locally update routes). There is no regard to how frequently or when route updates are desired. Routes, though, will always be available to transmitted packets on the network [45].

On-Demand routing techniques must wait for a route to be requested before it will form, causing a node to delay packet transmission until the route has been established (imposing a packet delay). As ad hoc networks are assumed to be energy constrained, this delay may be acceptable as the signaling overhead and power consumption is reduced due to the decrease in route information propagation [45]. A more detailed description of various ad hoc routing protocols is given in Chapter 4.

2.3 Effects of Signal Propagation

Transmitted signals must ultimately travel through environments which may produce various degradations. Based on the distance traveled along a given path from the transmitter to the receiver, the signal encounters free-space loss. Due to the signal's interaction with environmental objects, it incurs some type of shadowing gain. The final factor of a signal's loss or gain deals with the combined makeup of every surviving ray of the signal at the reception point. This multipath interference is due to many signals combining with varying phases and signal strengths. In general, therefore, the net gain or loss to a signal can be thought of as having contributions from all three of these different components.

2.3.1 Gain

The gain due to the distance traveled along a given path, or G_{distance}, is the gain of the signal due to the distance between the receiver and the transmitter. Based on various environmental factors, the magnitude by which the signal degrades due to distance may change. In general, the free space loss tends to follow an inverse-exponential law as follows:

$$G_{\text{distance}} = \frac{1}{d^n} \tag{2.1}$$

In this case, d is the distance while n is the exponential decay factor. The decay factor varies along with the environmental surroundings, and is generally between 1.2 and 6. The decay factor also depends on the distance. If the value of d is less than unity, for instance, the gain may reach a value which is greater than 1, which cannot occur realistically due to free-space attenuation. At this point, we can assume that the decay factor reduces to zero. Thus, for close distances, there is assumed to be no loss in the channel due to free space [28].

The gain of the signal due to objects impeding the path of a signal is given as the shadowing loss, or G_{shadow}. A ray of the signal may have to pass around or through different environmental structures, experiencing a shadowing gain or loss. The path of the ray can be divided into subregions, with each having a corresponding shadowing gain component. Each of these gains, then, can be combined in a multiplicative fashion to determine the net shadowing gain. The net associated shadowing gain can be calculated as:

$$G_{\text{shadow}} = 10^{(0.1)p} \tag{2.2}$$

where p is the shadowing constant, or the actual gain, in dB. This gain is spatially related to the receiver, but only in considering the path covered by the impending ray from the transmitter; it does not depend on the net distance traveled. For any arbitrary situation, a certain transmitter/receiver position will have a shadowing constant associated with it. In areas with urban characteristics, for example, the attenuation would be greater than if the area was rural and open.

Finally, the gain of the signal due to the combination of many signals from the transmitter to the receiver is given as the multipath interference gain, or $G_{\text{multipath}}$. In many cases, the direct line-of-sight path cannot be achieved. Reception is then via rays which have reflected off of other objects. At the reception point, each ray will have a different amplitude and phase, whereas the carrier frequency can be assumed to be constant. Under constructive interference, the rays may combine to give a stronger signal. Alternatively, under destructive interference, the net signal may be grossly attenuated. Like the shadowing loss, the multipath interference is not related to the distance between the two objects. In fact, the gain due to multipath interference can radically change as either the receiver or the transmitter move on the order of a wavelength. Because of this, the multipath gain is usually approximated by using a statistical model in which values are recomputed every wavelength of movement. These statistics usually come in the form of Gaussian random variables which, when considered in the in-phase and quadrature components, translate into Rician or Rayleigh random variables.

2.3.2 Analytical Methods

A more general description of the received signal is represented as a convolution between the transmitted signal $s(t)$, and some transfer function, $h(t,\tau)$. Specifically,

$$r(t) = \int s(t-\tau)h(t,\tau)d\tau \tag{2.3}$$

where the transfer function may be time variant. From a generalized point of view, this transfer function reflects all the aspects of the attenuation of a signal, as well as its interaction with the environment.

One method of analysis for the transfer function is to focus on the rays of the signal. These rays are formed in various ways, such as diffraction, scattering, reflection, and absorption. Thus, each ray travels a different distance, passes through different objects, and arrives with different amplitudes and phases, encompassing every aspect of signal degradation. Because every environmental configuration is different, a universal model is unfeasible. Similarly, analytical models are only attainable for specific geometries.

Nevertheless, a brief discussion on the properties of the transfer function is possible. First, we assume for simplicity that the transfer function is time homogeneous. A direct path signal delay can be defined as:

$$\tau_0 = \frac{1}{c}\left|\overline{x_t} - \overline{x_r}\right| \tag{2.4}$$

which is basically the ratio of the distance between the receiver and the transmitter to the speed of light. The line of sight path, though, is not the only path over which the signal might travel to the receiver. Other rays may travel over longer, possibly more complicated, paths, incurring a larger delay in reaching the receiver. If the net time for the *i*th ray to reach the receiver is given as τ_i, the additional delay is given as:

$$\Delta\tau_i = \tau_i - \tau_0 \tag{2.5}$$

Now, the transfer function, $h(\tau)$, can be seen as the amplitude response of the incident waves on the receiver from the transmitter. For any given ray, a different time delay and attenuation due to its specific path will have occurred. Thus, the transfer function can be specified as:

$$h(\tau) = \sum_i a_i e^{j\theta_i} \delta(t - \tau_0 - \Delta\tau_i) \tag{2.6}$$

Here, a_i represents the amplitude of the *i*th ray, and θ_i represents its phase. As each ray is discrete, the delta function associates the amplitude and phase with a specific incoming wave. The transfer function now represents a train of delta functions with different amplitudes and phases. If a continuum of rays is assumed, the value of $\Delta\tau_i$ approaches zero, and the transfer function becomes continuous.

2.3.3 Extension to Statistical Analysis

The previous discussion unfortunately does not provide ease of implementation. Specifically, determining the transfer function at all points would involve a calculation of all amplitudes and phases associated with each incoming ray. In early treatments, experimentation has provided accurate results for various geometries, involving extensive testing and measuring to describe the layout of the environment. Recent methods have attempted to use computer simulations to mimic the hands-on experiments. Ultimately, a reliable model is difficult to base on mathematical axioms, suggesting the use of empirical data.

It is in this spirit that we propose a statistical method of computation in Appendix A. To define the statistics for the entire system would be difficult, so each component of the signal attenuation is investigated individually. While the free-space gain is a deterministic value, the shadowing and multipath interference components can be assumed to be defined by certain statistics. This method has many advantages over deterministic models. For instance, little hands-on testing is required beyond that which is needed to define the statistics. Also, simulations can be executed in a short time, as opposed to extensive, deterministic computer simulations. Finally, any changes to the environmental parameters, such as object positions and statistical surface layouts, can be made easily without repeated testing.

For the purpose of the work here, the model presented in Appendix A was simulated in PARSEC [20] as a separate entity. This functionality keeps track of the locations of all the sensor nodes, as well as the environmental parameters, allowing transmitted messages to experience realistic signal degradation. As the model presented here does not require the a distinct set of parameters for each separate node pair, it easily allows the inclusion of hundreds to thousands of nodes.

Chapter 3

EAR: an Energy Efficient MAC Protocol for Mobile Nodes in HANETs

After a link level architecture and routing trees have been set up within the stationary sensor network, it is ready to support mobility. Mobile sensors are beneficial in that they provide topological variability, possibly acting as an information bridge where a stationary sensor has failed, and further the overall functionality of the network, providing the ability to spatially extend sensing operations. These mobile sensors can be in the form of robotic units or security personnel. For example, military personnel can move through a sensor field extracting data and other relevant information. Robotic sensors, on the other hand, can be deployed in areas where further required environmental sampling may not be possible by the static nodes due to constraints on sensing range. Furthermore, by allowing mobile nodes to move throughout the network, localized information extraction and network instruction is possible.

As the mobile nodes are able to possibly return to an energy reservoir to replenish its reserves, or to replace battery packs, it can be assumed that they do not share the same constraints as their stationary counterparts. Furthermore, as the entire network is possibly engaged in its own sensing tasks and network operations, the relatively few mobile nodes must operate protocols transparently to the protocols governing the operation of the stationary network. This suggests the need for novel protocol development to control the link level, or MAC (Medium Access Control) level, interaction between mobile nodes and the stationary network.

Section 3.1 describes the MAC level issues associated with networks that support mobile nodes, including MANETs and cellular networks, with an extension to HANETs. Section 3.2 develops the EAR protocol to form, and maintain, connectivity to the stationary network with a low drain on the stationary node's energy reserves. This section also describes some of the benefits associated with densely populated HANETs. Section 3.3 gives simulation results, while section 3.4 concludes the chapter.

3.1 Medium Access Issues in Networks with Mobility

Incorporating mobile nodes within a stationary network environment involves the design of protocols which perform the dual-edged task of supporting link level connections and routing information to the mobile nodes. Here, we will focus on the MAC (medium access control) protocols associated with mobile nodes along with the corresponding signaling and resource allocation required to form, maintain, and sever connections to the stationary network. The development of these protocols must take into account various issues which are common to all MAC level protocol designs. In particular, these design issues fall into the categories of Quality of Service (QoS), resource costs, and distributive functionality.

Maintaining a high QoS while attempting to keep resource usage down is the primary tradeoff associated with protocol design. In the situations of interest here, mobile nodes will require some guaranteed level of service, usually measured in connection quality (received SNR) and spatial connection reliability (outage probability). A high connection quality will increase the probability of signal reception, while offering the possibility to increase throughput via sophisticated signal modulation techniques. Furthermore, a low outage probability will allow the mobile nodes to spend more time connected to the network. Maintaining a high QoS, though, usually involves a higher resource usage cost. A higher connection quality, for example, is easily achieved by increasing the transmission power at the nodes, resulting in higher energy depletions. Similarly, increased signaling can help resolve reception errors which lead to outages (shifting power in power-controlled schemes, or detecting low signal quality), again with the increased energy cost at the radio level. While striving to achieve appropriate QoS and resource usage requirements, the protocols design may also attempt to move towards a more robust and distributed functionality.

The methods by which a given network manifestation supports MAC level connections for their mobile nodes, therefore, are dictated by the requirements of the overall network. It is conceivable that various networks will incorporate vastly different protocols to support mobile nodes based on the connectivity goals and available resources. We proceed to examine two prominent networks which support mobility, Mobile Ad-Hoc Networks (MANETs) and Cellular Networks, along with the features of Hybrid Ad-Hoc Wireless Sensor Networks (HANETs) which motivate the development of new MAC level protocols.

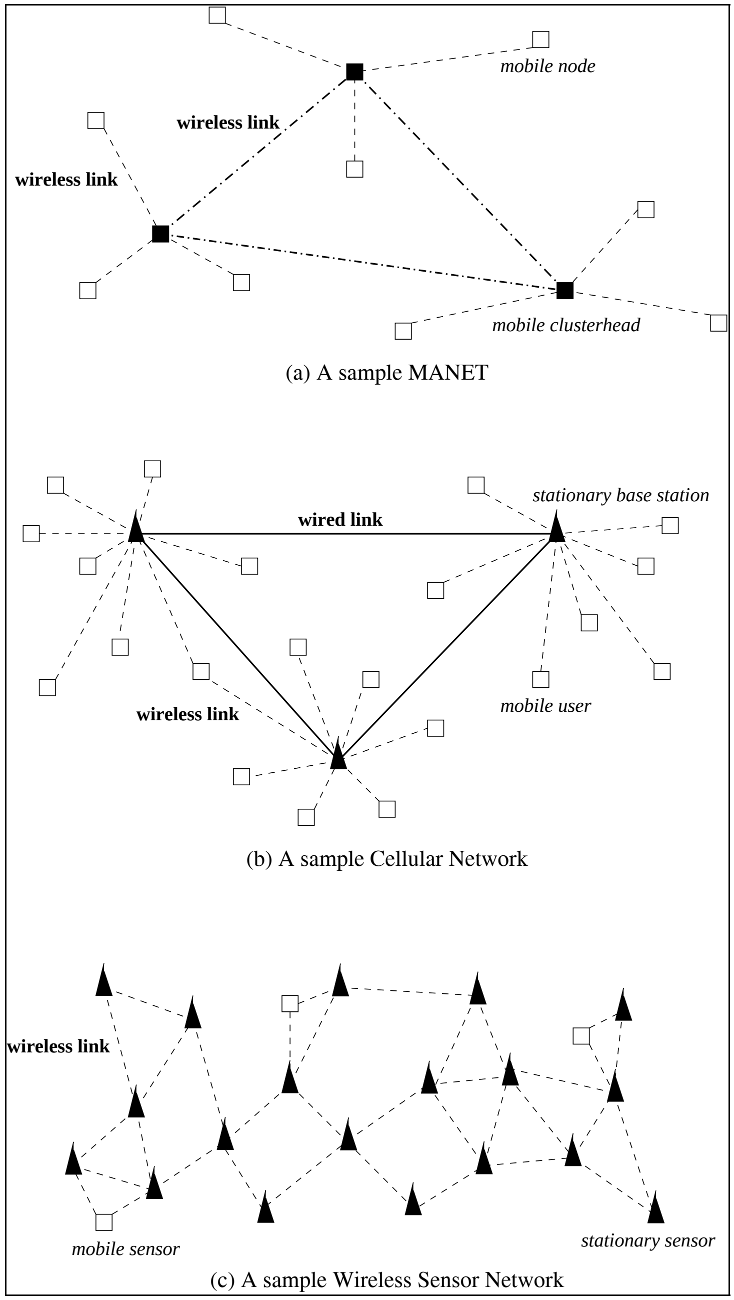
3.1.1 MANETs

A MANET is a peer-to-peer network which usually comprises tens to hundreds of communicating nodes that are able to cover ranges of up to hundreds of meters. MANETs have been studied in various forms, including Mobile ATM [47, 59] and Mobile Packet Radio [15, 32]. Protocol design for MANETs attempts to optimize the provided QoS. That is, the network is designed to provide good throughput/delay characteristics in the face of high node mobility. The mobility management issues, in particular, have been classically oriented toward routing issues within the network. Since the network consists solely of mobile nodes, the tasks of routing and mobility within the MANET are generally handled jointly. One way proposed to handle these networks is to group the mobile nodes into small clusters, electing a clusterhead to which to route information in a local neighborhood [22, 38, 49]. The group of clusterheads in the entire network in turn forms a subnetwork. Information is then routed through this subnetwork. As mobile nodes move from one area to the next, they may decide to register within a new cluster and continue operation as usual.

Although the nodes are portable battery-powered devices, energy consumption in this system is of secondary importance. Each device is always attached to a person, and presumably the depleted battery will be replaced when needed (the same way batteries are changed on laptops). Thus, the primary goal of MANETs is to maintain network connectivity and organization, with secondary importance given to preservation of energy reserves. Figure 3.1(a) gives an example of a clustered MANET network configuration.

3.1.2 Cellular Networks

A cellular network is a vast network consisting of both stationary and mobile nodes. The stationary nodes, or base stations, are connected in a subnetwork with a wired backbone, forming a fixed infrastructure. Mobile nodes greatly outnumber stationary nodes (tens to hundreds of mobiles per base station), which are usually situated quite sparsely. The base stations' locations are pre-planned so as to cover a large region with little overlap from cell to cell. The wired backbone that the stationary nodes form facilitates routing, as the wireless channel is avoided. Consequently, it is only the single hop from a mobile node to the stationary base station that needs to be considered. Thus, mobility management is primarily considered here from the point of view of forming connections with the base station offering the highest signal quality. As the mobile users travel from the vicinity of one base station to the next, the desired connection is simply updated using any of many handoff techniques, and communication continues as normal [24, 55, 57]. In a general scenario, base stations will continually transmit pilot signals throughout the network. Mobile nodes will respond to these pilot signals by immediately forming a connection with the transmitting base station. Power control methods ensure that the mobile node will usually connect to one base station, with multiple connections allowed in the cell overlap regions.



(a) A sample MANET

(b) A sample Cellular Network

(c) A sample Wireless Sensor Network

Figure 3. 1 Various wireless networks

Since the base stations are assumed to have a large energy reservoir, they take on much of the responsibility for mobile management (setting up new routes to the mobile nodes, informing mobile nodes of handoffs, etc.). Alternatively, the mobile nodes are assumed to be battery-operated with the ability of recharging or replacing depleted energy supplies. Thus, energy preservation is of secondary importance in protocols designed to operate in cellular environments. The primary goal here is to provide low outage probability to the mobile nodes, along with high bandwidth efficiency, enabling each base station to support more mobile users. Figure 3.1(b) depicts a sample cellular network.

3.1.3 HANETs

The manifestation of mobility in HANETs is significantly different than that of MANETs and Cellular networks. While mobile nodes are dominant in either of the other two network types, HANETs are comprised primarily of stationary nodes forming a network with its own protocols (i.e. multihop routing, distributed bootup) and operational goals. The relatively few mobile nodes are later introduced into the network requiring connectivity support. The stationary nodes in HANETs are assumed to have limited energy supplies, requiring low-power protocols to be developed to increase the lifetime of the network.

Using protocols developed for MANETs and cellular networks may prove to be difficult, due to the network goals and resource constraints. Similar to MANETs, HANETs will primarily utilize the wireless channel, suggesting that it may be possible to incorporate their protocols by simply assuming various nodes have no mobility. HANETs, though, do not offer their nodes the ability to transmit at large distances (multihop routing is the dominant form of information transfer). Furthermore, energy conservation is the primary concern, as it is important to increase network lifetime in the face of added mobile functionality. Thus, protocols designed for HANETs will move away from those operational in MANETs. Cellular networks provide an interesting comparison to HANETs. While the cellular network architecture involves many mobiles per stationary base station, HANETs are comprised of many stationary nodes per mobile node. Also, in stationary networks, the mobile nodes are assumed to be energy limited (with the option of rechargeable batteries), while in HANETs it is the stationary nodes which require aggressive energy saving protocols. But, utilizing handoff techniques introduced for cellular networks will suggest frequent handoffs for low-power transmissions by stationary nodes, as smaller cell sizes will imply a higher stationary node density and more frequent cell transitions.

The primary design goal of developing new MAC level protocols for HANETs is to provide connectivity to the mobile sensors in the face of properties which distinguish these HANETs from more conventional networks which support mobility. In particular, there are far fewer mobile nodes present in the network relative to stationary nodes, and the stationary nodes are considered to have high energy constraints relative to the mobile nodes.

Fewer mobile nodes: Having a low number of mobile nodes implies that the interaction between any particular stationary node and a mobile node will be rare. It is important, therefore, that the mobility support protocols run transparently over the stationary network protocols, thereby avoiding the disruption of static network functionality. Also, it is not feasible to provide each mobile node with the locations of the stationary nodes, as the nodes may initially be randomly dispersed within the network area.

Energy constraints: As the stationary nodes are energy constrained, the protocols designed for maintaining mobile connectivity must drain as little energy from the stationary nodes as possible. In highly dense sensor fields, frequent handoffs using so-called "relative handoffs," where the mobile user will connect to the stationary node offering the best signal quality, must be avoided. Furthermore, it may not be attractive to provide the stationary node with control in connection maintenance schemes.

Both properties mentioned above bring about the same question: Is stationary node control inefficient? Giving these nodes control implies unnecessary polling with the purpose of searching for mobile nodes which may never be present, due to their few numbers. Also, this signaling imposes a regular drain on energy reserves at the stationary nodes. We propose a solution to this problem in which the HANET adopts a mobile-centric MAC protocol solution. The mobile node will assume much of the responsibility of connection maintenance, while utilizing the ongoing functionality of the stationary network's MAC layer protocols.

3.2 EAR: Eavesdrop and Register

The EAR algorithm (Eavesdrop and Register) allows mobile sensors to maintain connectivity to a wireless stationary sensor network, while preventing extensive energy consumption at the stationary nodes. It accomplishes this by allowing the mobile nodes to remain inconspicuous to, but to continuously monitor, the stationary network, initiating handshaking procedures only when desired. Before the algorithm is presented, though, it is important to examine the assumptions made as to the properties of the stationary network. We assume that the sensors are randomly distributed, perhaps with no ability as to determine location and proximity to other nodes. Furthermore, these nodes have limited battery supplies which we assume are not replenished when consumed. The stationary sensors are operating a slotted TDMA-type frame structure, possibly utilizing frequency hopped spread spectrum techniques from slot to slot, with synchronization taking place on a link to link basis. At some point during its frame structure, the stationary node enters a "searching" phase, which consists of a polling signal which is used to invite other stationary nodes into the network (assumed to be at a known frequency), followed by a set of slots within which another stationary node may respond. For example, if a large set of nodes are depleted of energy supplies, a new set can be distributed throughout the network. This "searching" functionality will allow the new nodes to achieve network connectivity.

To allow the mobile MAC protocol to operate transparently to that of the stationary network, we propose that the mobile nodes use the features of the existing MAC protocol. In particular, to avoid specialized pilot signals, or polling, the mobile sensors can simply listen for these "searching" messages, which act as a trigger for the EAR protocol. If we were to adopt a scheme similar to conventional cellular networks, these messages can be used to initiate handoffs to achieve connections to the stationary nodes with the highest signal quality (SNR). But, in a dense network, as we have with

HANETs, handoffs may not always be necessary. In one case, a mobile may set up one connection, and pass by multiple stationary nodes before swapping connections to avoid the signaling overhead associated with handoffs. Also, a mobile node may tend to remain in a confined area, perhaps requiring only one connection to be maintained. But, to attain the ability to remain selective regarding connection options, the mobile node should be able to acknowledge many stationary neighbors, as well as connections.

The mobile node will have the ability to maintain a registry, which will contain all information regarding surrounding nodes such as their ID number, connection status, received signal quality, and transmitted signal power. This information can be inferred from the stationary network's regularly transmitted pilot signal. Upon eavesdropping this pilot signal, the mobile node can register the stationary node as a neighbor, with the EAR protocol dictating connection formation, with both statuses depending on received signal quality. Using this registry, the mobile node can establish a sense of "movement" for itself, allowing it to initiate its own handoff procedures. The stationary node, on the other hand, will only be responsible for receiving invitation or disconnection messages from the mobile node, as well as data communications.

The signal quality experience at the receiving mobile node can be quite high. This is achieved by forcing the connection threshold SNR to be much higher than the required SNR. The mobile nodes can tolerate this because the stationary nodes are assumed to be in close proximity, due to their high density and low transmission ranges, providing more stationary connection options for the mobile nodes. By forcing a higher overall SNR, the EAR protocol can allow the mobile nodes to be more selective regarding possible connections. Also, acknowledgement messages can be avoided by assuming probable message receptions and employing timeouts.

3.2.1 Messages

The EAR protocol employs a three message scheme. If the pilot signal associated with the stationary MAC protocol is also assumed as a message, though, four messages are used. These are as follows:

Broadcast Invitation: This message is regularly transmitted by the stationary nodes at some point within their frame structure. There is no guarantee as to when a general stationary node will transmit a BI message. Thus, the mobile nodes will need to employ radio control techniques allowing the radios to be powered on or off at any given time. Through this message, a mobile node can extract the sending node's ID, the received signal quality and the transmitted power.

Mobile Invite: Upon reception of BI messages, the mobile node forms a registry of surrounding nodes. At some time, the mobile node will wish to initiate a connection request, accomplished by transmitting an MI message to the corresponding stationary node. In some cases, the mobile node may need to have connection priority over all other stationary nodes. This is accomplished by having the mobile node transmit the MI message in the slot immediately following a BI message.

Mobile Response: Upon receipt of the MI message, the stationary node will determine if a connection is either possible or desired. In situations where energy reserves at the stationary node are low or no channels are available for communication, the stationary node will opt to decline the connection request. In this case, either the MR is sent as a declination, or the MI message is simply dropped, allowing the mobile node to deduce that no connection was desired. If a connection is possible, the stationary node will respond with a MR message indicating a desire to accept the connection. This message can also be avoided by allowing the stationary node to simply transmit data in place of the MR message.

Mobile Disconnect: After connecting to a stationary node, the mobile node will continue to monitor the received SNR of the corresponding stationary node. At some point, either the received SNR will drop below a required disconnect threshold, or the stationary node will disappear (due to a deep fading effect). In either case, the mobile node will transmit this MD message to inform the stationary node that a connection is no longer desired. Upon receipt of this message, the stationary node will simply remove the mobile node from its connection list.

3.2.2 Algorithmic Details

The EAR protocol continually operates in three phases, a registration phase, a connection phase, and a disconnection phase.

Registration: Upon reception of the BI message, the mobile node will attempt to enter the corresponding node's information into its registry as simply a neighbor. Continuous registration of multiple stationary nodes will allow the mobile node to determine approximate relative proximity to the various stationary nodes. As the registry size is finite, it is possible that a node is detected which has not been registered, for which no empty registration slot exists. In this case, the mobile node will determine the node with the worst received signal quality among those to which connections have not been formed. If the new node offers a higher signal quality, the mobile node will assume it is closer, and replace the inferior node entry with this new one.

If a BI message is received from a stationary node which has already been registered, the mobile node will simply update the required information (signal quality, update time) in the registry. Further updates of received signal quality offers the mobile node to employ open loop power control techniques for communicating with stationary nodes.

Connection: As the registry contains the received signal quality of all prospective connection candidates, or neighbors, in the mobile node's vicinity, it contains all the required information for determining connection selection. If the mobile node is allowed to connect to multiple stationary nodes, a maximum number of connections allowed to be experienced by the mobile node can be set. When a connection is required, the mobile node will simply scan the registry for the best possible connection, determined by crossing a received signal quality threshold. This value can initially be set quite high (10-15 dB), with the possibility of reduction if no stationary link exists with a high enough signal quality.

The desire to form connections arises in one of two fashions. In the first case, if the maximum allowable connection threshold has not been reached, the reception of BI messages from simple neighbors may trigger a received SNR to rise above the connection threshold, resulting in a connection request. In the second case, as a current connection's received SNR drops below the disconnect threshold (and the maximum number of connections has been reached), the mobile node will be allowed to disconnect from the current connection and form a new one. Note, though, that once connections have been formed, and the mobile is connected to the maximum allowable number of stationary nodes, a new connection request cannot be made until a current connection fails (handoff dependent on absolute SNR). Figure 3.2 depicts a sample mobile activity in the network.

A field exists in the registry to specify the connection status of the stationary nodes. This entry can be either NOT_CONNECT, CONNECT, or PENDING. Upon registration of a neighbor, the mobile will enter a status of NOT_CONNECT for the stationary node. When a connection is requested, the status will shift to PENDING, awaiting a response from the stationary node. Once the stationary node has accepted, the mobile node will update the status to CONNECT. For the purposes of eliminating nodes due to new neighbor candidates and determining if a mobile node has achieved the maximum connection allowance, a PENDING status is treated as a CONNECT status. This avoids the possibility of eliminating a node which may accept a connection or surpassing the connection allowance threshold. The PENDING status is used, though, to facilitate in dropping possible connections due to timeouts.

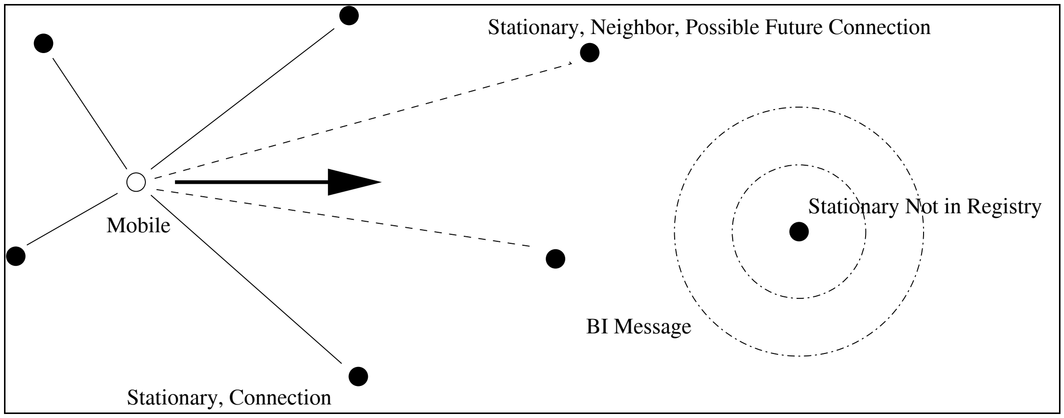


Figure 3. 2 A sample mobile node activity

To request a connection, the mobile node will send an MI message to the stationary node. Concurrently, the status of the corresponding stationary node is upgraded to PENDING. Along with the connection request, the mobile node will offer the stationary node a selection of possible channels for communications (slot pairs and frequencies). As the two nodes may not know the others' frame structure, slot pairs are offered as offsets from the current slot of communications. Furthermore, as the two nodes may not have synchronized frames, slot buffers can be used, allowing for half the slot length to be used for communications. These slots associated with this stationary node will not be offered to another node until dropped as possible communications channels. Upon receipt of the MI message, the stationary node will determine its desire to participate. If a connection is possible, the stationary node registers the mobile node, while selecting a subset of communications channels from the offered set, and responding to the mobile node during the next selected slot with an MR message. At this point, all of the channels not selected by the stationary node are uncommitted, and can be used to invite other nodes. Also, the status of the stationary node is upgraded from PENDING to CONNECT.

This MR message represents the only specialized message introduced to the stationary node by the EAR protocol. It is even possible to avoid this message. For example, if the stationary node has data which can be transmitted to the mobile node, then it will simply transmit this data in place of the MR message in the next selected slot. If a connection is not desired, the stationary node is not required to transmit an MR message, and it simply discards the MI message from the mobile node. The mobile node, therefore, will wait a predetermined time, and then drop the connection possibility and reduce the status of the stationary node from PENDING back to NOT_CONNECT.

Disconnection: After forming a connection with a stationary node, the mobile will continue to update the corresponding entry in the registry by receiving data and BI messages. Either two events will occur causing a disconnection: (a) the received SNR will drop below the disconnect threshold, or (b) the stationary node will "disappear." The disappearance of a node occurs when no data or BI message has been received within a preset time limit, usually due to node failure (depleted energy reserves or malfunction) or deep fades due to shadowing or multipath interference. In such cases, the mobile node will transmit an MD message (at the highest power level to accommodate for disappearing nodes) to inform the stationary node of the failed connection.

3.2.3 Mobile Radio Control

We assume that the mobile sensors are able to control their radios via the MAC protocol. In particular, during any slot, the radio may be switched in receiving mode, transmitting mode, or simply powered off. While in receiving or transmitting mode, the radio is also able to control its frequency selection. As the radio dissipates energy whenever it is operating, it is important to incorporate intelligent radio control for energy efficient operations at the mobile nodes. The EAR protocol assigns a status to every slot within the frame structure, consisting of the power status, corresponding stationary node ID, and communications frequency (or hopping pattern).

Initially, the frame consists of a "blank slate," with slots being filled as connections are formed. For each connection, at least one slot pair is set aside for connection maintenance requirements, during which the node couplet will be forced to exchange channel update information in the presence of no data. During unfilled slots, the mobile node is free to eavesdrop the channel to search for further connections. It is feasible to allow the mobile to continuously monitor the channel during these empty slots, but this would cause a drain on energy reserves. Thus, one proposed scheme is to only probe the channel continuously when no connections are present, and then simply reduce the probing rate as a function of the number of connections. In such cases, the mobile node can avoid probing for BI messages when its maximum connection limit has been reached. Chapter 5 investigates more aggressive radio control techniques.

3.2.4 Relative vs. Absolute Handoffs

Handoffs occur as a mobile node moves from the vicinity of one stationary node to the next. To ensure the best connection quality, relative handoff schemes are traditionally employed. After connecting to a stationary node, a mobile will continue to monitor the channel for pilot signals corresponding to potentially better connection candidates. When a new signal is received, the mobile node is allowed to connect to two stationary nodes (soft-handoff), until such time when one of the two signals overcomes the other (in terms of SNR) by a preset level. This guarantees the mobile node will always experience the best possible link. For sparse networks, where the coverage areas of the stationary nodes overlap only at the fringes of the cell, this is an acceptable solution. But, for densely populated HANETs, which are energy constrained, this introduces a high level of signaling which can be avoided.

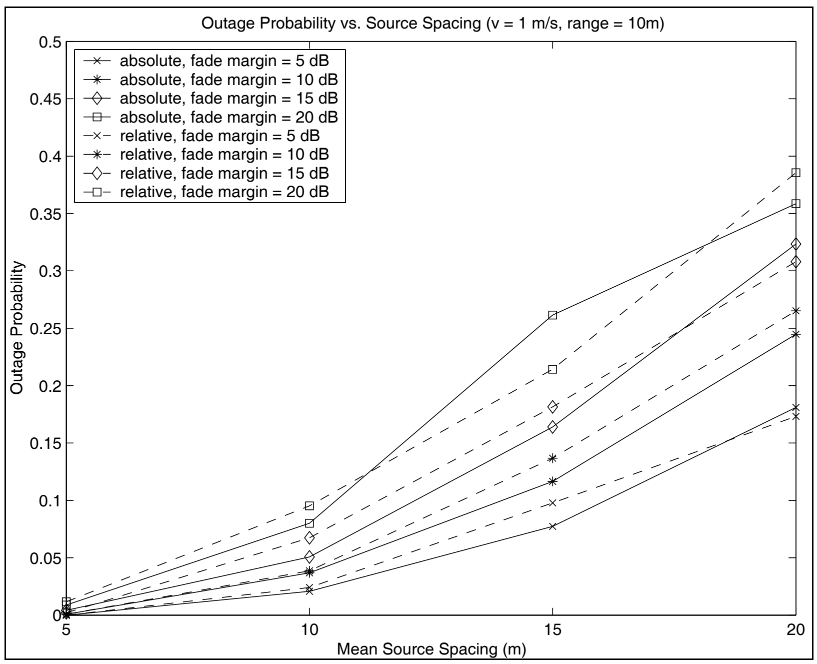


Figure 3. 3 Outage probability for mean source spacing using relative and absolute handoffs

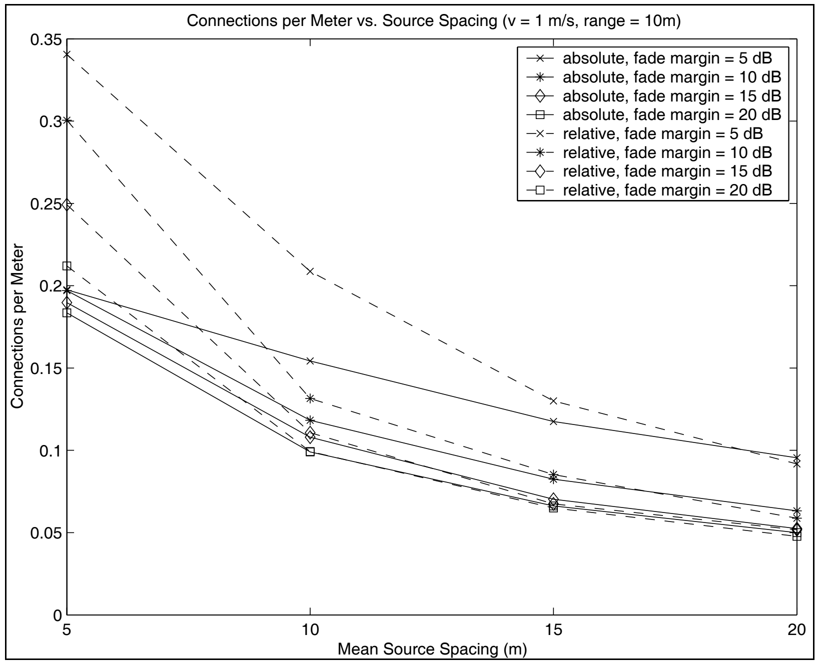


Figure 3. 4 Connection signaling overhead using relative and absolute handoffs

As the HANETs are comprised of stationary nodes which are connected via a wireless medium, the coverage area of each node must extend to the next stationary nodes, creating a large overlap region. Thus, it is possible that the mobile node will constantly be changing stationary connections, even though high quality connections may already exist. Thus, the EAR algorithm uses an absolute handoff criteria. Here, each connection is treated separately, as opposed to comparatively for relative schemes. A connection is formed as the signal quality surpasses a connection threshold, and severed as the signal quality descends below the disconnection threshold. Once a connection is formed, though, it will not be replaced until a disconnection takes place. To ensure reliable communications, the disconnection threshold is usually set higher than the SNR required for communications.

Simulations were performed to determine the effectiveness of absolute handoff versus relative handoff. Sensor nodes were randomly dispersed in a field according to a Poisson distribution, with the mean of the sensor node separation varying from 5 meters (highly dense) to 20 meters (sparse). The communications range of each node was set at 10 meters, with the mobile node velocity at 1 meter per second. The fade margin was varied from 5 to 20 dB. Figure 3.3 shows the outage performance of both handoff schemes. Note that there is no significant difference in outage performance when migrating from relative (dashed) to absolute handoff (solid). Figure 3.4 gives the signaling overhead in terms of the number of connections per meter that the nodes must experience to maintain connectivity. For dense networks, there is a significant reduction in the signaling overhead by using absolute handoffs, approaching a savings of 1 connection per 7 meters for low fade margins.

By increasing the connect/disconnect thresholds, it is possible to control link selection, thereby controlling the quality of the average connection experienced by the mobile node. A high connection threshold will increase the average received SNR, at the cost of a higher outage probability (see next section), as the probability of attaining higher signal qualities is low. Increasing the disconnection threshold will guarantee that no connections experience a low signal quality, but it will also increase the signaling overhead as the mobile will need to initiate tradeoffs more often to replace connections.

3.2.5 Timeouts and Acknowledgement Avoidance

By assuming a high density environment, the EAR protocol is able to bypass the traditional use of acknowledgement messages (ACKs) by employing a set of timeouts and enabling ACK avoidance. Three types of timeouts are constantly running at the mobile nodes:

Stationary Disappear Timeout: Whether a stationary node has attained the status of a connection or a simple neighbor, the mobile node will monitor the instances of the corresponding registry updates. A disappear timeout threshold is set to a number of frame lengths. If the stationary node has not sent information warranting a registry update with this set number of frame lengths, it is dropped from the registry list. If it had already been promoted to a connection, an MD message will be transmitted at a higher power level to accommodate for the phenomenon causing the disappearance.

Mobile Invite Timeout: After an MI message is sent by the mobile node to request a connection from a stationary node, it will begin to wait for a response. A threshold for the MI timeout is again set to a number of frame lengths. If a response is not received by the stationary node within the allotted time, the mobile assumes that either (a) the stationary node did not receive the message (transmission or node failure) or (b) the stationary node is not able to make the connection (lack of available channels or low energy reserves). In either case, the mobile node will drop the present request and initiate new connection requests.

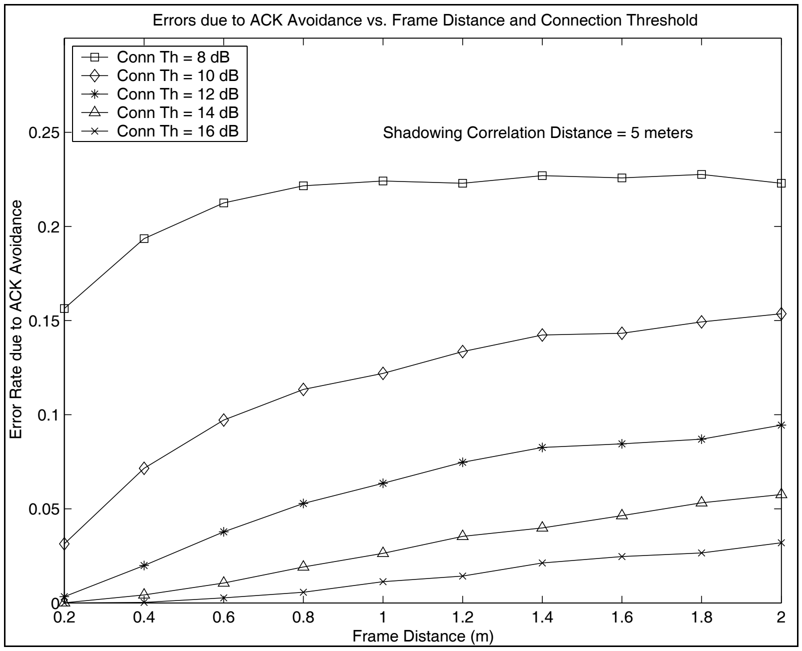


Figure 3. 5 Message errors due to ACK avoidance in a 5m correlated shadowing environment

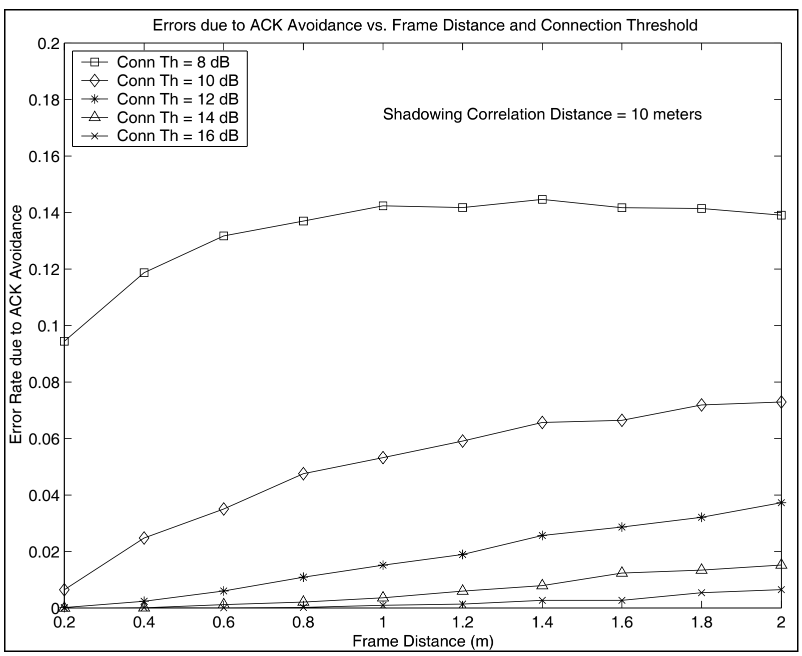


Figure 3. 6 Message errors due to ACK avoidance in a 10m correlated shadowing environment

Disconnect Timeout: After a connection has formed, the mobile node will continue to monitor the received SNR from the stationary node, initiating the disconnect timeout if the signal level falls below the disconnect threshold. To prevent unnecessary disconnections due to momentary signal fluctuations, the signal is required to fall below the disconnect threshold for a preset number of frame lengths (the disconnect timeout) before an MD message is sent.

Beyond the use of timeouts, though, the messaging scheme used by the EAR protocol does not require the use of any ACK messages. This is achieved by relying on the high density of the network, and the ability to raise the connection threshold required by the mobile nodes. Because the received signal quality must be higher than the connection threshold (possibly significantly higher than the required SNR for communications), the transmitted messages will almost surely be received. To illustrate this point, simulations were performed in which a mobile node responded to a BI signal of varying connection thresholds (dB). The error rate signifies the probability that a message sent after the reception of the BI signal is not received. Tests were run for varying frame distances (the distance the mobile moves in one frame length). Figures 3.5 and 3.6 give results for 5 meter and 10 meter shadowing correlation distances [17]. Note that as the connection threshold increases, in either case, the error rate significantly drops. Furthermore, in densely distributed sensor networks, a higher connection threshold can be supported, due to the number of available stationary neighbors with high connection qualities, without an increase in outage probability (see Figure 3.8).

3.3 Results

A simulation testbed for the EAR protocol was implemented in PARSEC [20]. In this simulation, each node is modeled as an *entity*, with separate functions written for the various protocol levels. Along with stationary MAC protocols [52] and multihop routing tree formation [14], the EAR protocol completes a protocol suite described in [52]. The simulation models the network lifetime, including initial node dispersion, bootup and link formation, routing formation, and mobility support (MAC and network level routing), for hundreds of nodes.

To model the radio propagation channel, a separate function was implemented as an entity which includes spatially correlated shadowing and path loss. Multipath interference was not considered, due to mitigation abilities using diversity and error control techniques [53]. The radio channel model designed for this testbed incorporated a stochastic shadowing grid technique, described in appendix A.

A network consisting of 100 nodes was used for this simulation, at a density of 0.4 nodes/m^2. Each stationary sensor was assumed to be using a fixed transmit power of 1 mW, while each mobile node was able to vary its transmit power to three fixed levels, 1 mW, 3 mW, and 10 mW. The frame length for the stationary nodes was 8 seconds, consisting of 40 slots per frame. The mobile nodes were allowed to function with frame lengths of 10 slots, allowing a factor of 4 increase in the number of slots offered for communications with stationary nodes. Each mobile node was allowed to register up to ten neighbors, with the ability of sustaining three connections at any given time. Each stationary node, on the other hand, was only allowed to form a connection with one mobile node. At the time of connection request, the mobile node was able to offer 5 slot pairs to the stationary node for communications (amounting to twenty options with a factor of 4 of slot reuse), of which the stationary node could choose up to 2 sets. During a connected state, it was assumed that the two nodes would transmit data during each of its assigned slots. The mobile node's velocity varied between 0 and 2 m/s.

At the radio channel level, the shadowing grid used had a gridpoint spacing of 1 meter, which is the approximate correlation distance, with shadowing values following a log-normal distribution (underlying Gaussian variable with a standard deviation of 8 dB). As the sensor nodes are located near ground level, a distance based power attenuation following a fourth power law was used. To accommodate for a fade margin at the transmitting stationary sensors, the required SNR to register a BI signal was -5 dB. The SNR threshold to form a connection to a stationary node was varied from 6 to 15 dB, while the allowable drop in signal quality (again measured in SNR) varied from 1 to 5 dB. The mobile node was allowed to monitor the channel during all slots which were not taken up for communications with other sensor nodes.

Figure 3.7 shows the average bit error rate for received packets for various connection thresholds and allowable drops in signal quality for a binary PAM modulation scheme. Note that as the connection threshold or the disconnection threshold is increased (the allowable signal drop is decreased), the bit error rate decreases. As mentioned earlier, for sparse networks, increasing the connection threshold will yield an increase in outage probability, due to the difficulty in forming new connections. Figure 3.8, though, shows that for dense networks, as used in HANETs, there is not an increase in outage probability for the connection thresholds of interest here. In fact, the outage probability primarily depends on the disconnection threshold, with higher thresholds yielding a higher outage probability. Thus, we can increase the connection threshold, along with decreasing the bit error rate, to ensure packet reception without an increase in outage probability.

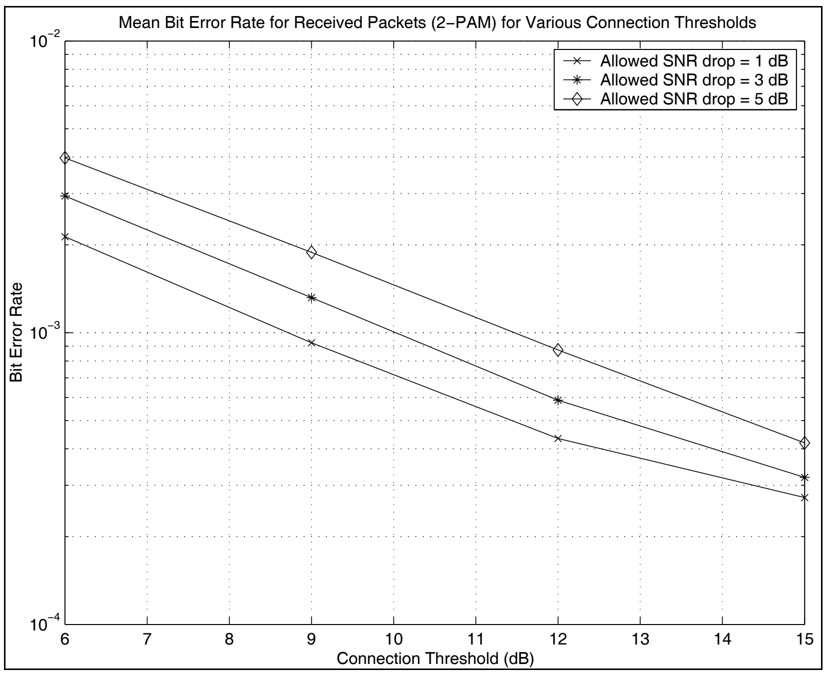


Figure 3. 7   Mean BER for received packets using EAR

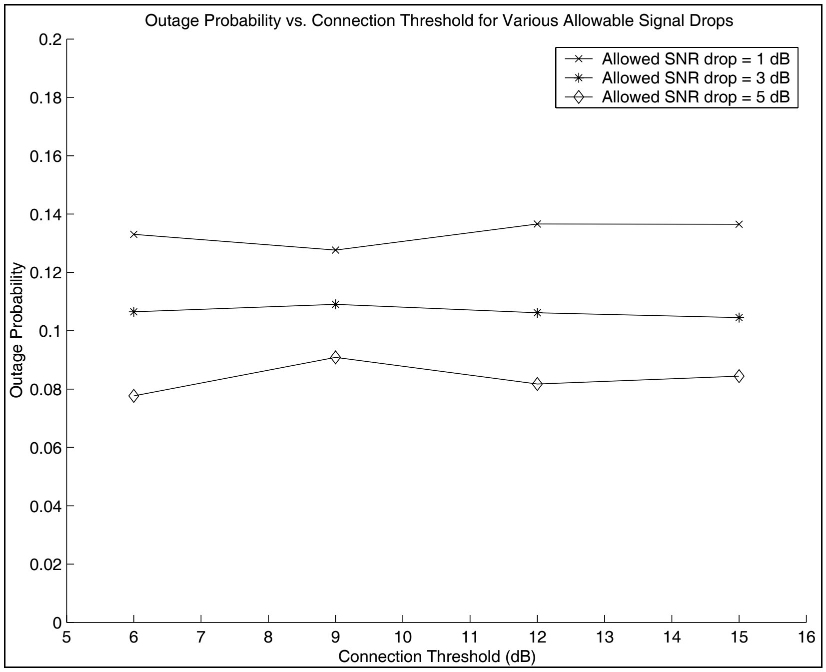


Figure 3. 8 Outage probability for mobile nodes using EAR

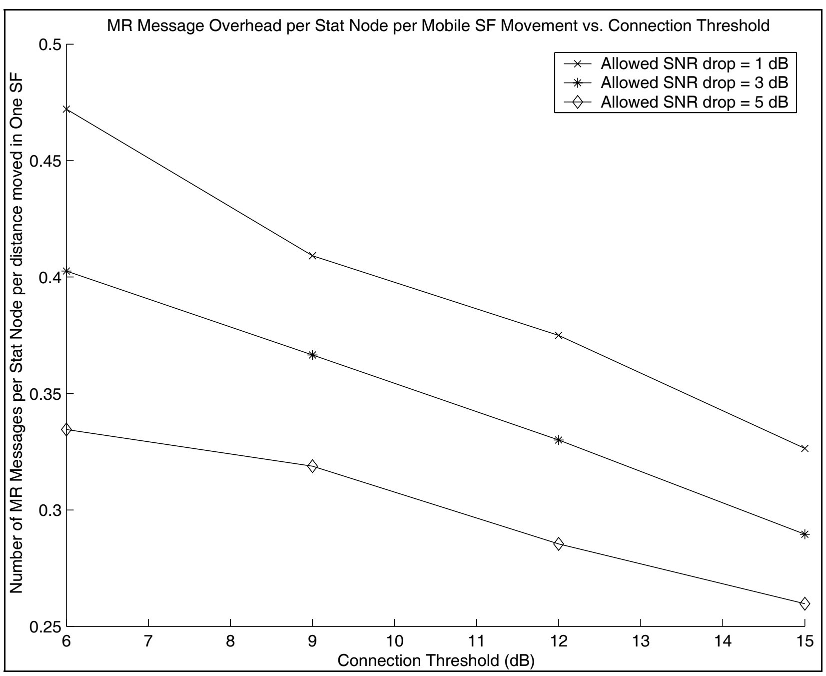


Figure 3. 9 Stationary node signaling overhead using EAR

Figures 3.9 and 3.10 depict the signaling overhead experienced at the stationary node and at all nodes, respectively. Again, the independent variables are the connection and disconnection thresholds. The signaling overhead has been normalized to the number of messages (MR for figure 3.9, all 3 for figure 3.10) per distance moved in one frame length (product of frame time and velocity). From these figures, it can be seen that of all of the signaling overhead experienced in the network, only about 20% is attributable to the stationary nodes. Figure 3.11 gives the throughput available to the mobile node in slot pairs (channels) per stationary frame. For low connection thresholds, it is more likely that the mobile node will form multiple connections, allowing an increase in the number of available channels offered to the mobile node in steady state operations.

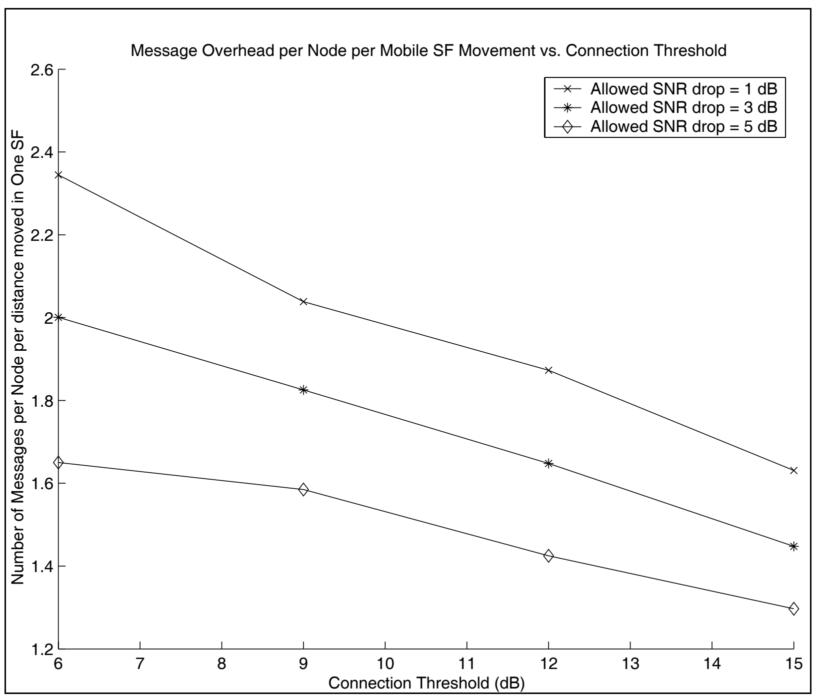


Figure 3. 10   Overall signaling overhead using EAR

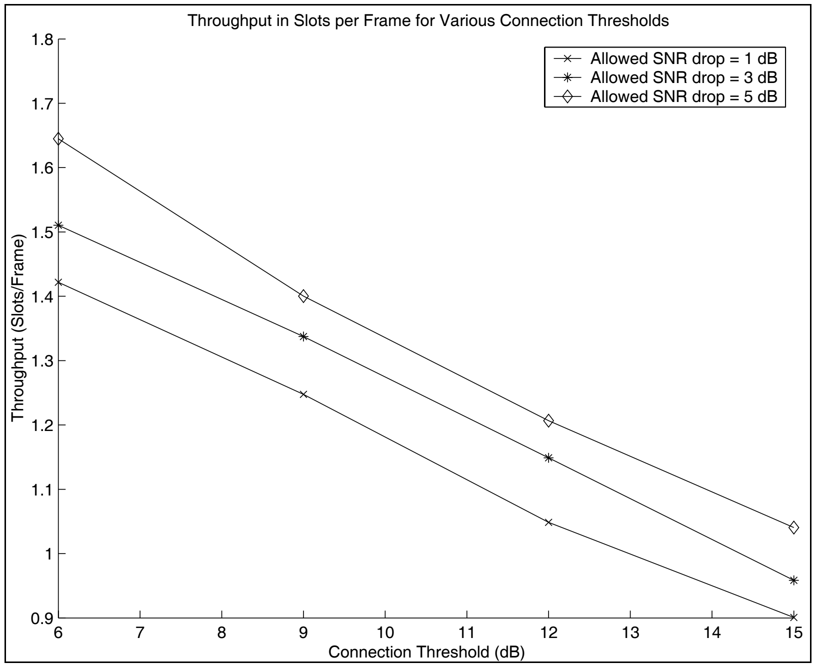


Figure 3. 11 Throughput using EAR

3.4 Conclusion

A novel concept of Hybrid Ad-Hoc Wireless Sensor Networks (HANETs) is presented, along with comparisons to conventional mobility-support networks such as MANETs and Cellular Networks. Due to constraints on energy resources, as well as a large available bandwidth and a low ratio of mobile nodes to stationary nodes, new protocols must be designed for HANETs which provide connection maintenance to mobile nodes.

The EAR protocol, or Eavesdrop and Register, is developed, which suggests a mobile-centric view of MAC level connection maintenance. Essentially, the mobile node utilizes the MAC level features of the stationary protocol, intervening when necessary to form connections. A 4 message scheme is used (including one message native to the stationary protocols), of which only 1 has been introduced for the stationary nodes. The mobile node will form a registry of neighbors, deciding on connection promotions as needed. The dense nature of the stationary sensor network suggests that (a) absolute signal levels should be used in the place of relative levels and (b) acknowledgement messages can be avoided without an increase in outage probability.

The protocol was tested in PARSEC, along side other network protocols, to determine the behavior of mobile MAC level connections within a sensor network. Results show an ability to maintain high signal quality in the face of mobility, as well as a nominal participation by the energy-constrained stationary nodes to maintain network connectivity for mobile nodes.

Chapter 4

MIR: an Intermediate Rerouting Protocol for Mobile Nodes in HANETs

As the mobile nodes are able to act as information sinks, as well as sources, provisions need to exist to allow packets to be routed to and from these mobile nodes across the stationary network. The nodes themselves, which are assumed to have constrained energy supplies, have a limited range, and are likely to be communicating via a multihop routing method. One shot communications, therefore, between a mobile node and the sink node (or any stationary node to the mobile node) is not possible. Mobile nodes must rely on the multihop mechanism of the stationary network to propagate information efficiently in the face of mobility. This suggests the need for novel protocol development to quickly set up routing paths throughout the network, supporting intelligent packet forwarding mechanisms, while adhering to quality of service constraints such as low signaling overhead (energy constraints), low packet delay, and high probability of packet reception.

Section 4.1 describes the Networking level issues with networks supporting mobility, including cellular networks and MANETs. An extension to the routing requirements of HANETs is discussed as the motivation for a new protocol development. Section 4.2 introduces the MIR protocol, as well the two prominent algorithms used to set up paths and route packets to and from mobile nodes. A comparison is also made with the routing methods used in a variety of networks. Section 4.4 gives simulation results compared to other packet forwarding mechanisms. Section 4.5 concludes the chapter.

4.1 Routing Issues in Networks with Mobility

The presence of mobile nodes in a network involves the design of protocols which are able to continually offer these nodes link level connections and network level routing abilities. We focus here on the networking protocols associated with mobile nodes, along with the corresponding signaling required to form routing paths, and the packet forwarding mechanism to transport packets to and from the network. Various protocols have been suggested for network level functionality, each incorporating the abilities, goals and requirements of the specified network. All of these protocols can be classified by their methods of route formation and packet forwarding. Routing in networks supporting mobility has traditionally assumed that the entire network is mobile, whereas only recently there has been motivation to combine the abilities of both mobile and stationary sensor nodes. Therefore, we will present the basic trends in networking protocol design, followed by the motivation for a novel protocol development for HANETs.

4.1.1 Trends in Ad Hoc Network Protocol Design

In setting up routes through the network, sensor nodes will act as routers, and will take either a proactive or reactive approach to path formation. In the proactive, or Table Driven, case, nodes will set up long term routing tables leading to each node in the network. To set up these tables, though, the network will need to flood connectivity information from node to node, propagating complete path specifications. Reactive routing, also known as On Demand routing or Source Initiated routing, will attempt to find the best path from the source to the destination immediately prior to message transmission. Although this method will not incur the extensive signaling overhead associated with routing table formation, it does impose a delay on each message transmission as paths are established. Networks are usually concerned with this tradeoff between signaling and delay, as the nodes are not assumed to be energy constrained. Thus, protocols are classified by their proactive or reactive nature. If energy reserves at the sensor nodes are taken into account, then, along with signaling to set up routes, nodes must be able to efficiently set up routes (low signaling to/from mobile nodes) and transmit packets throughout the network (low delay, guaranteed delivery).

Common methods of packet transmission, as well as the occasional signaling associated with path formation, include flooding, broadcasting, and link-to-link unicasting. Flooding occurs when a single message is propagated to every node within a local area, or the global network. This method can be used to find routes (send a request packet to all nodes, with the intended recipient informing the source of the optimal path), or to bypass routing requirements completely (flood the network with the message in the hope that it will eventually lead to the destination node). Even if flooding is avoided, many mobile networks will transmit information via broadcasting. This, though, requires sustained transmissions and multiple receptions, thus incurring a high energy cost. As an alternative to each of these, link-to-link unicasting allows a transmitter to send packets to the intended receiver via preset channels (time slots, frequencies, codes), presenting an efficient alternative to both broadcasting and flooding, but also requiring specific MAC level attributes (orthogonality in channel selection and formation).

One of the most basic networking protocols supporting mobile nodes is a cellular network [24, 57]. Mobile issues in these networks are generally handled at the MAC level, as the routing is facilitated by the stationary backbone network formed by the static base stations. The MAC layer protocol will assure that the mobile user will always remain one hop away from the stationary network, thus facilitating routing issues. The initial route to the mobile node is established as the mobile device powers up, followed by a continuous series of exchanged messages between the mobile node and the base station, maintaining local routing updates. In a sense, routing in cellular networks is closely related to routing in MANETs using clusterhead based routing. In fact, [33] suggests that mobiles can also act as routers within a cell, setting up multihop paths from a mobile node to the base station, which is acting as a clusterhead node.

MANETs, as opposed to cellular networks, handle mobility management by focusing on routing issues at the mobile level, rather that MAC issues. Methods to route packets from node to node within MANETs fall into either a hierarchical or a flat topology. Clustering is an example of a hierarchical topology [10, 15, 50], where each node member is within k-hops from any clusterhead. A subset of all mobile nodes is chosen as a set of clusterheads (set participation may require regular reconfigurations), which are responsible for routing packets from one region to another. MAC level connections to the clusterhead, therefore, act as single hop routes to a clusterhead node (for 1-hop clusters).

Flat topologies, on the other hand, are required to set up routes based on either a proactive or a reactive basis. Table Driven routing protocols, such as DSDV [39] and WRP [34], require each mobile node to maintain and update routing tables to all nodes within the network. Information, such as number of hops to a specified node, as well as various paths leading to that node, must be continuously updated at all nodes, occasionally requiring acknowledgement messages. As these methods tend to impose a decrease in network capacity due to a high level of signaling, various reactive routing algorithms have been suggested. In protocols such as AODV [40] and DSR [23], the source node will initiate a path discovery process, by flooding a control message (small packet), when a message (large packet) is to be sent. The intended recipient will then reply to the sending node with the optimal path information. As an extension to DSR, TORA [37] and LAR [25] propose schemes which allow for localized rerouting, thus avoiding flooding a control message due to slight changes in the network topology. Other approaches avoiding the use of routing tables are ABR [54] and SSR [11], each of which continually monitors some metric of the link-level connections to suggest stable routes. In ABR, each node will periodically transmit a beacon, receptions of which will suggest some level of mobility (frequent receptions imply low mobility). SSR, on the other hand, monitors the signal quality achieved for each connection.

Recent protocols have suggested a combined use of both proactive and reactive methods. ZRP [41] has proposed that table driven methods can be used to route packets from zone to zone within the network (within k-hops of a destination), with inter-zone routing performed on a demand-only basis. In Terminode Routing [4], multiple anchored paths are set up as general waypoints. The remote routing method (TRR) is responsible for routing information to known friend nodes, continuously attempting to route to nodes which reduce the distance to the recipient node. The local routing method (TLR) is similar to that of ZRP.

4.1.2 HANETs: The Introduction of a Stationary Wireless Backbone

The coexistence of a large number of stationary nodes along with relatively few mobile nodes separates the network architecture of HANETs from that of MANETs. Prior to the inclusion of mobile nodes, the HANET will establish a fully operational stationary network, including a network level multihop routing scheme. Thus, it is possible that as mobile nodes are included into the network, the stationary network can act as a wireless routing backbone. As the stationary network in HANETs is assumed to be energy constrained, though, protocols that are used for routing must efficiently determine paths and forward mobile-related information packets. As all of the protocols discussed above assume some level of mobility in all of their constituent nodes, they do not treat energy as a prime resource, and thus assign it a lower priority in favor of low delay and guaranteed delivery. This suggests the development of a new protocol which (a) reduces the signaling overhead in routing to and from mobile nodes by utilizing the abilities of the stationary network, and (b) provides a low delay and a high likelihood of packet reception.

Like the Terminode Routing method and ZRP, the combination of a wireless stationary backbone supporting mobile sensors will include both proactive and reactive route formation. The stationary network, after some path discovery phase, will have established proactive routes, perhaps multiple routes, from any sensor node the sink node. One method, discussed in [52], assumes that each sensor node has a list of its own neighbors, as well as an identification of the link which leads to the sink node. Thus, messages originating from any point in the network that are intended for the sink node are forwarded via the stationary backbone network. As mobile nodes are introduced, routes to and from the sink node must be established, with localized rerouting to compensate for mobility to reduce the energy consumption at the stationary nodes, suggesting a reactive localized routing method. Furthermore, flooding should be avoided in both the path organization phase and the packet forwarding phase.

4.1.3 Three Types of Routing

In HANETs, the primary direction of message forwarding is towards the sink node. The mobile nodes may act as either information sources or sinks, possibly providing network instruction or extracting localized information. This suggests three types of routing classes:

Mobile to Sink: Data generated at the mobile sensors may need to be routed to the sink node. As the mobile node is connected to the stationary network via one or more links, it will simply transmit the data to one connection, allowing the stationary network to forward it to the sink. This trivial case, therefore, is solved by MAC level connection formation.

Sink to Mobile: Packets originating from the sink node will need to find their way to the mobile node. To avoid flooding, it will be necessary to form paths leading down to the mobile nodes. At this point, two issues arise. First, these paths must be efficiently updated, as setting up new routes to mobiles as they move will be costly. Second, after a packet is sent, it is possible that a mobile node forms new connections, causing a lost packet. Thus, intelligent forwarding schemes must be used to reroute the packet, assuring delivery with low delay. Figure 4.1 shows an example of this problem as node B must correctly route the packet to node Z.

Mobile to Mobile: It may be necessary for one mobile node to transfer data to another. Utilizing the above two classes of routing, it is possible to forward a packet from the first mobile to the sink, and then to the second mobile. This process is unattractive as it will impose an unnecessary delay on the packet and an increased probability of failed delivery. Furthermore, there is increased energy consumption at each node which is involved in the propagation. Thus, a protocol must be designed which allows for packets to be rerouted to mobile nodes from other mobiles as required. Figure 4.1 demonstrates this ability as node X intends to route a packet to node Y.

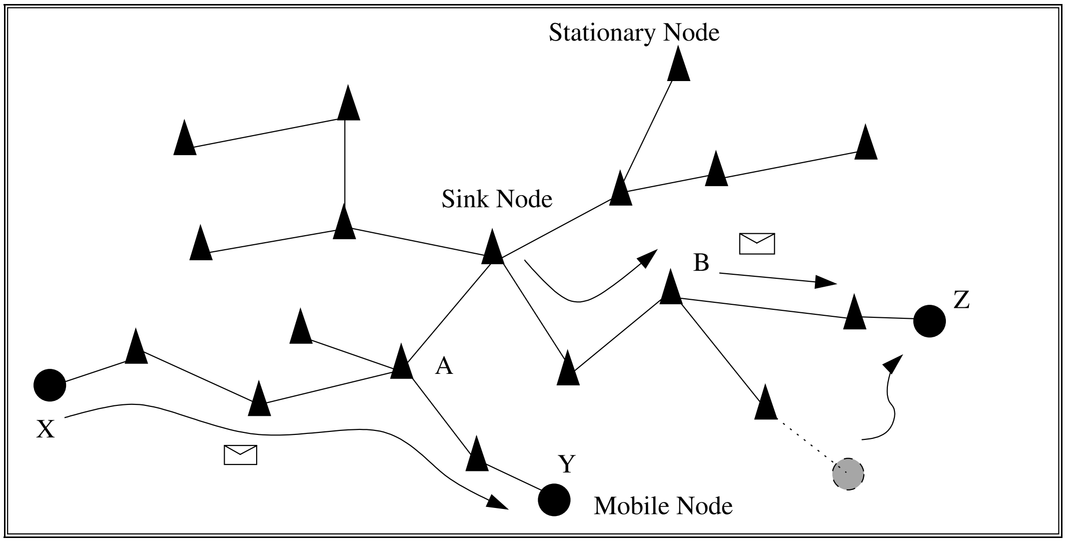


Figure 4. 1 Routing instances for mobile nodes

Along with attempting to fulfill the above mentioned routing requirements, the protocol design must take into account some fundamental issues associated with energy conservation on the stationary network. First and foremost, flooding must be avoided in path organization. To avoid unnecessary signaling, it may be possible to utilize the routing abilities already established by the stationary network to set up reverse routes to the mobile nodes. Furthermore, even as routes from the sink node to the mobile nodes may have formed, it should be possible to locally adapt the routing path without informing all nodes along the path to the sink node.

4.2 MIR: Mobile Intermediate Rerouting

The MIR protocol (Mobile Intermediate Rerouting) allows mobile nodes to maintain uplink (to the sink node) and downlink (to the mobile nodes) paths across the wireless stationary sensor network. Also, this protocol will reroute packets already present on the network in a loop-free manner to guarantee routing to the receiving node if it exists on the network. The uplink routes are assumed to be handled by the stationary routing protocol, with the downlink routes being set up with as little energy consumption at the stationary nodes as possible. Before the algorithm is developed, though, it is important to examine some of the features assumed by the stationary portion of the HANET. The network is randomly distributed, with perhaps no ability to determine relative location or proximity to other nodes. Each stationary node is energy constrained, requiring the development of efficient protocols to prolong the lifetime of the network. During the bootup phase, links are formed throughout the network, with each sensor communicating to its neighbors via orthogonal channels (TDMA slots and/or frequency assignments) to limit localized channel interference. Network level routes are then formed, which we will assume are multihop in nature. It is possible that another scheme is used, such as direct sensor to sink communications, but this will generally require a higher power output. The relatively few mobile nodes which are added to the network after the stationary network formation will interact at the link level by some MAC protocol, such as the EAR protocol presented in Chapter 3. The choice of the protocol is immaterial, though, as all that is required is that some type of handshaking occurs to inform the stationary node that a connection to the mobile node has formed. This connection information will serve as a trigger for the MIR protocol. A mobile node is assumed to be able to form multiple connections to the stationary network. It is feasible, therefore, that each connection formed to the stationary network will imply a downlink route to the mobile node.

The motivation for the MIR protocol arises by examining the routing trees that will potentially be formed from the sink node to the mobile nodes. In particular, we examine the multiple routing trees that may form to any single mobile node. At this point, we can assume that the mobile node is not moving, and that the stationary network employs some method which can be used to set up a reverse route to the mobile node. Each connection that a mobile node forms will translate into one route. In the uplink direction, it is likely that these multiple routes will combine at some point, merging into one single path. For packets being sent in the downlink direction, up until this merged point, only one path exists, implying that all nodes above this combination point need no knowledge of the multiple paths. Now, we can extend this notion to a mobile which has changed location. If the reverse route updating is accomplished quickly, then the mobile node can establish new routes upon forming a new connection. Since the stationary nodes involved in the handoff process are likely to have a high spatial correlation, it is again likely that the new routing path formed will intersect the previous path at some point (possibly meeting at the sink node). Intuitively, nodes located in the uplink direction need not be informed as to the routing change. It may be possible, then, to update routing paths on only the nodes located in the downlink from a potential combination point. Furthermore, when routing packets to the mobile node, it may also be possible to redirect misrouted packets to the proper destinations rather than dropping the packet or rerouting it back to the sink node. This combination of path updating and packet routing suggests two levels of protocols for MIR routing.

4.2.1 MIR: Path Update Algorithm

As mentioned earlier, uplink paths are resolved by forming a link level connection from the mobile node to the stationary network (first hop), and allowing the stationary network to handle the further routing (transparent to the mobile). The downlink path formation, though, will rely on the stationary network multihop capabilities. At the point of connection, the mobile node will cause the connected stationary node to originate a short control message (MOBILE_CONTROL) to be sent to the sink node. Each node along this uplink route will now have knowledge as to the downlink node which corresponds to this mobile node. This process is repeated each time the mobile node forms a stationary network connection. Each time a reverse downlink path is created, it will definitely intersect the previous path at some point (possibly at the sink node). To alleviate the signaling due to the uplink transmission of the control message, the stationary node which is also the path intersection point will be able to drop the control packet, as the path will be identical to the uplink nodes. Similarly, if the mobile node forms multiple connections, with paths leading through the same stationary node, that stationary node need not forward the control message. Also, this stationary node will now have a choice of paths on which to route packets down to the mobile node.

The important metrics in route formation are corresponding mobile node ID, route update time, and hop count to the mobile node. Thus, as the MC propagates upward, each stationary node, including the initiating node, will extract this information from the packet, with each node incrementing the hop count upon forwarding. The stationary nodes each will maintain a table (proactively) of the corresponding mobile ID, the hop count, the route update time, the link status (ACTIVE now, possibly DEAD later), and the corresponding outgoing link ID (the sender of the MC message). Upon reception of an MC message, the stationary node will check to see whether a link to this corresponding mobile node already exists, which would imply that all nodes on the uplink are aware of the downlink route. If the corresponding mobile node does not have any existing outgoing link from this stationary node, all of the information carried in the MC message is extracted, and the MC message is forwarded towards the sink node.

If an outgoing link to the corresponding mobile node exists, the stationary node will record the time update information, and will need to see whether a change has occurred in the hop count or link status. Only three conditions will warrant the MC message to be propagated up to the next link towards the sink node:

New Hop Count: As will be discussed below in the routing section, a stationary may have multiple choices as to the outgoing link to the corresponding mobile node. As the stationary node receiving the MC message may be a link along a path which will join another path in the uplink direction, it is important to propagate all changes upwards to improve decision making abilities. Thus, if the hop count has changed along this path, the stationary node will forward the MC message.

No More ACTIVE Links: When a mobile node disconnects from a stationary node, the stationary node will generate an MC message informing all nodes on the uplink that this link status is DEAD. It is possible, though, that the message is not propagated back to the sink, but is dropped intermediately due to the ACTIVE link criteria. If a node receives a DEAD link message, but has another possible link to the corresponding mobile node, it will drop this MC message, as the uplink nodes will still be able to route to the mobile node via this stationary node. If, on the other hand, upon dropping this link, the stationary node has no outbound link to the mobile node, the MC message is forwarded, and the link status is updated to DEAD after a short delay (which allows a new route to possibly form). Note that a DEAD link is still maintained as a best possible guess to the last known location of the mobile node, and, as such, only the most recent DEAD link status should be allowed to exist, with the others being dropped.

DEAD to ACTIVE Status: If a mobile node has revived a path through a stationary node which is NOT corresponding to a DEAD link, the MC message will propagate as it is assumed to be forming a new path. Also, if the corresponding outgoing link has a current DEAD status, the status is then changed to ACTIVE, followed by MC propagation.

Using the algorithm as described above will not only effectively form reverse routes to the mobile nodes from the sink node, but will also attempt to adapt to mobility induced topological changes locally, thus avoiding the signaling overhead corresponding to MC messages. Figure 4.2 demonstrates the local adaptation ability of the MIR protocol.

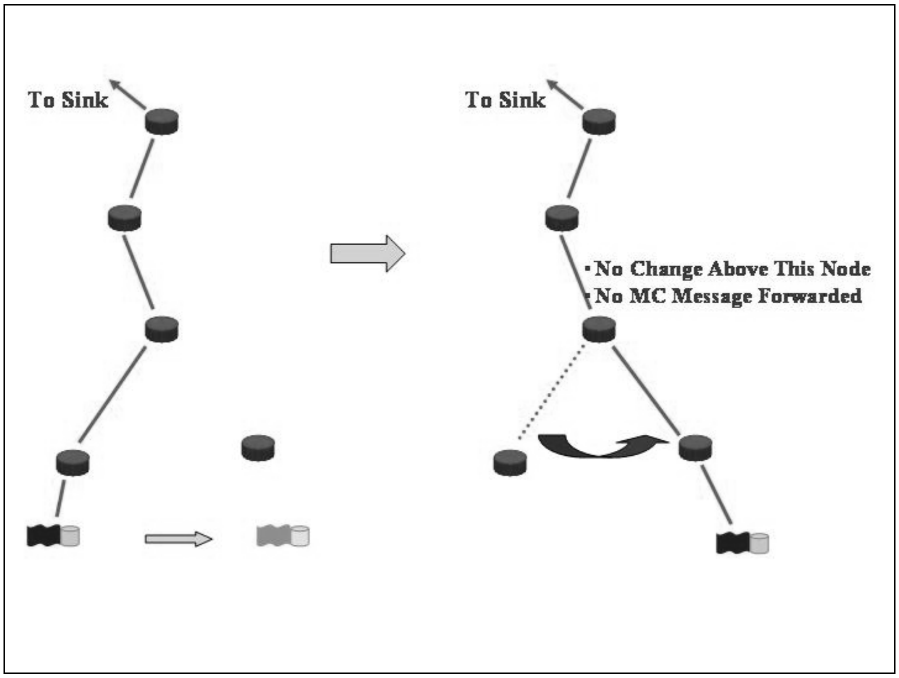


Figure 4. 2 Example of the MIR: Path Update algorithm

4.2.2 MIR: Routing Algorithm

The routing portion of the MIR protocol depends on the formation and maintenance abilities of the Path Update algorithm. Thus, if there is no data available for the mobile node to transmit to the connected stationary node, the mobile node may need to transmit an MC message as a beacon to continuously confirm the local route. In such cases of simple route confirmation, the receiving stationary node will simply drop the MC message according to the Path Update portion of the protocol. Here packets are re-routed to their intended destinations if a mobile has changed location. To illustrate this idea, intermediate packet rerouting will first be compared to other common techniques such as back-to-sink rerouting and drop-on-error method. Following this, we will present a detailed description of the Routing Algorithm.

In Back-to-Sink rerouting algorithms, a packet which as been misrouted will be forwarded back to the sink node to determine the next best possible route. This technique is quite stable, as all reverse routes will inevitably be rooted at the sink node. Unfortunately, this method will typically involve a higher overhead (due to multiple forwarding hops) and a higher packet delay, suggesting applications requiring guaranteed delivery specifications, such as data transfer. The Drop-on Error technique, on the other hand, avoids this signaling and delay overhead by simply dropping the packet when misrouted. This method guarantees low delay when a packet is successfully delivered, but does not guarantee delivery, thus lending itself to applications such as real-time voice and video transmissions. Intermediate rerouting, as used in the MIR protocol, attempts to maintain a low signaling and delay overhead in the face of node mobility, while achieving a high packet delivery rate. Packets which have been misrouted are forwarded on the uplink to the previous stationary node to determine a possible alternate route. If none is found, the next uplink stationary node will receive the packet, repeating the process. This continues until the packet reaches the sink node. If no alternate route is discovered by this point, the packet is finally dropped, and the mobile node is assumed to have disappeared. As very few protocols have attempted to utilize the wireless stationary routing backbones found in HANETs, the above routing methods provides an alternate method of comparison to other protocols.

Each mobile oriented data packet will contain a header field, including the direction of the packet. This directional information is used to determine the best possible route for the data packets destined for mobile nodes. Three direction classes exist: UP, BROKEN, and DOWN. UP is used to designate that the packet is moving in the uplink direction, and that the packet is not being routed upwards due to a broken link. A BROKEN direction, on the other hand, signifies that the packet is moving upwards, and that the previous link is not a valid path to the intended mobile node. Packets with a DOWN direction specification are moving downlink along a reverse route to the intended mobile node. The details for route selection are as follows:

DOWN: Stationary nodes receiving packets in the DOWN direction will attempt to determine if an ACTIVE route exists down to the mobile node. If one or more exists, the stationary node will select a path based on some decision criterion involving such metrics as the minimum hop count and the most recent route update time. If no ACTIVE link exists, the stationary node will determine if a DEAD link exists, signifying the best possible route selection (if the DEAD link does not correspond directly to the corresponding mobile node). If neither of the above cases hold, then the packet is assumed to be misrouted, and is forwarded to the stationary node on the uplink after a delay time (with a BROKEN direction specification). This delay time will allow the mobile node to possibly reform a new path upon connection to a new stationary node.

BROKEN: Packets being received in the BROKEN direction will originate from nodes which were incorrectly assumed to be hops along the route leading to the intended mobile recipient. The corresponding link to this mobile node, therefore, will need to be eliminated from the stationary node's routing table (not changed to DEAD). At this point, the stationary node will again check to determine if an alternate route exists to the receiving node. As in the above case, if one or more alternate ACTIVE links exists, the packet is rerouted downward based on some determination criteria (direction header changed to DOWN). If no other ACTIVE routes exist, and the status of the chosen route was actually a DEAD link, then no other DEAD routes will exist either (due to the DEAD route formation functionality described in the Path Update algorithm). In the case where no alternative routes exist, the packet is propagated upwards, continuing to carry the BROKEN direction header. If a packet reaches the sink node with a BROKEN header and no alternate routes exists, then the packet is dropped, and mobile is assumed to have disappeared.

UP: A stationary node receiving a packet in the UP direction will assume that this message has originated from a mobile node, intended for either the sink node or another mobile node (based on the recipient node ID contained in the packet header). The treatment of this packet is very similar to the BROKEN case, with the exception that the immediate sending nodes are not eliminated from the routing tables. Assuming that that packet is destined for another mobile node, the MIR protocol will attempt to avoid forwarding this packet back to the sink node prior to downlink routing. Thus, each stationary node receiving this packet will determine whether a route exists to the receiving mobile node (using the idea of ACTIVE/DEAD link determination discussed above). If a downlink option exists, the packet is rerouted, with a DOWN direction header; otherwise, it is simply propagated upwards to the sink node.

Using the intermediate rerouting features described above for the MIR: Routing Algorithm, will maintain a low signaling overhead and delay characteristic associated with routing to mobile nodes. Furthermore, this will accomplished with a low packet dropping rate yielding a higher probability of packet reception.

4.3 Results

A simulation testbed for the MIR protocol was implemented in MATLAB and PARSEC [20]. The PARSEC version was modeled to complete a protocol suite, as described in [52], which also includes a stationary link-level bootup algorithm for a randomly dispersed sensor field, as well as a network layer multihop routing tree formation protocol. Each node was modeled as a separate entity, with separate function within the entity written for each layer. The entire simulation was able to model the lifetime of a network consisting of hundreds of nodes, from node deployment, bootup and link formation, routing formation, and mobility support (including MAC, Network, and Radio level functionality).

A separate entity was created to act as the radio propagation channel. This model, which incorporates spatially correlated shadowing and path loss, is described in Appendix A as incorporating a stochastic shadowing grid technique with variance preserving interpolation. Multipath interference was not considered for this simulation, due to the mitigation abilities using diversity and error control techniques [53].

The performance of the MIR protocol is highly dependent on the features of the nodal makeup of the HANET, including stationary network density and mobile sensor speed. In low density situations, it may not be possible for the stationary network to form routing trees from any node to the sink node. In other words, if the stationary routing protocol breaks down, the MIR protocol has no wireless backbone to depend on. Under situations of high node mobility, a MAC level algorithm must quickly set up new links, allowing for robust localized path updating. If new links do not form fast enough, the packets will consistently be dropped. If the mobile's velocity is arbitrarily increased, even in the presence of a sufficient connection update protocol, message will continue to be rerouted, incurring a high delay if the message is finally delivered.

Various HANET configurations were tested, including stationary network densities ranging from 50 to 100 nodes per 2500 m^2. The mobile node velocity was normalized to the distance traveled in one frame structure (corresponding to the next available communications slot), by incorporating the temporal length of the frame. This frame distance was varied between 0.5 to 2 meters. An average range of 10 meters was assumed for all nodes, taking into account shadowing effects using a lognormal shadowing model with an 8 dB variance for the underlying Gaussian random variable. Using these specifications, along with a required received SNR given a modulation scheme, the stochastic received SNR can be determined by the radio propagation expression derived in chapter 5. To ensure connectivity to the stationary network implied that a reverse route from the sink node to the mobile node exists, tests were run on networks which were 100% connected. Mobile nodes were allowed to form up to 3 connections to the stationary network.

Simulations were run to determine the packet dropping rate and the received packet delay for each of the routing methods described above, Drop-at-Error, Reroute to Sink, and Intermediate Rerouting (MIR). As Rerouting to Sink will provide similar results as MIR when compared to the packet dropping rate, we compare our protocol to the Drop-at-Error method. Figure 4.3 shows the behavior of both protocols as the density of the network is increased. As the density is increased, the Drop-at-Error method tends to provide worse performance, whereas the MIR protocol actually improves. In Figure 4.4, the performance of the Drop-at-Error protocol is shown to degrade substantially in the face of high mobility and/or frame length. The MIR protocol, on the other hand, only shows slight performance degradation, with less than a 5% dropping rate for packets. This is due to the fact that higher network densities, as well as high mobility, imply that many handoffs will occur at the mobile link level. For the MIR protocol, though, the packets are simply rerouted as new links form. Thus, for higher densities, and highly spatially correlated stationary node connections, the rerouting provides robust packet reception.

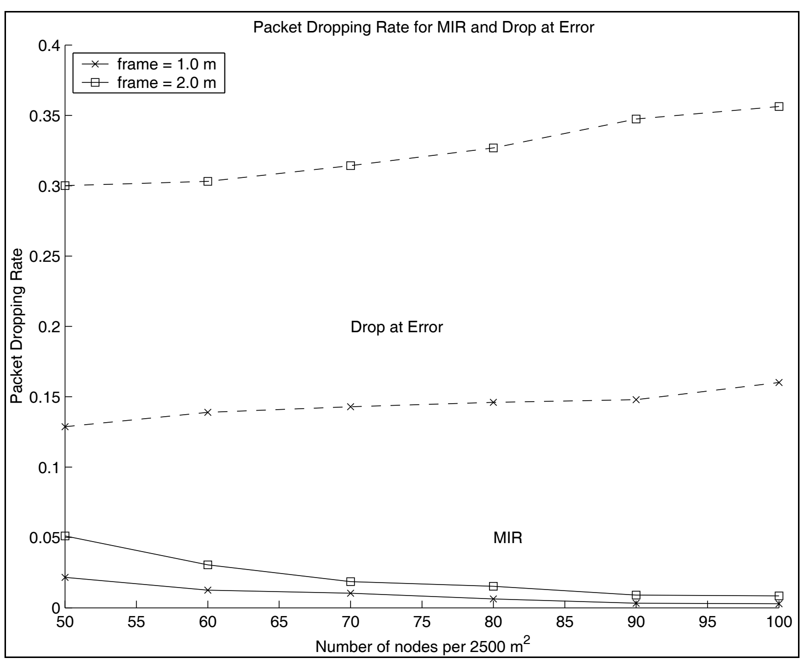


Figure 4. 3   Packet dropping rate for MIR vs. Drop-at-Error

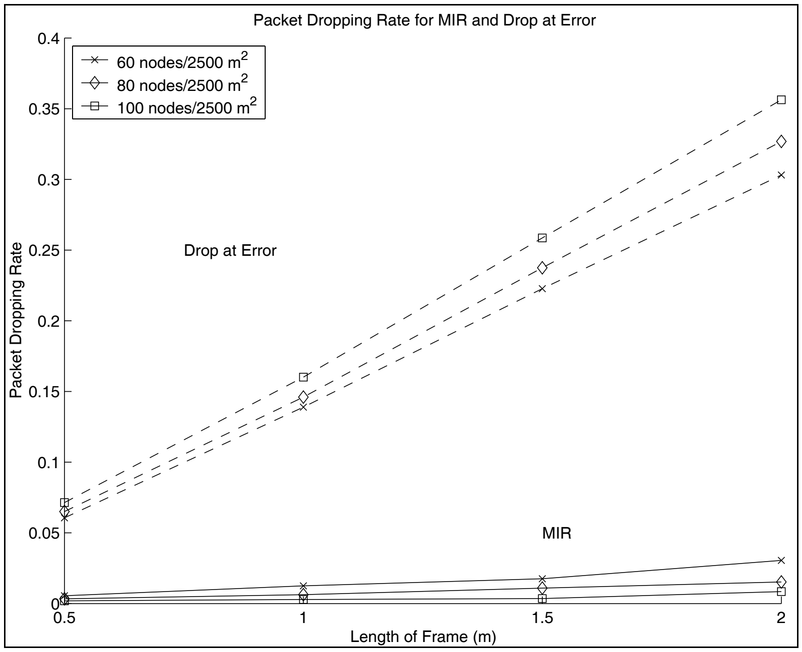


Figure 4. 4 Packet dropping rate for MIR vs. Drop-at-Error

The MIR protocol was compared to the Reroute Back-to-Sink routing method to determine the comparison in relation to packet delay, which is directly related to signaling overhead (multiple hops). Figure 4.5 shows that both routing methods tend to impose a higher signaling overhead to the stationary nodes as the velocity of the mobile node is increased. In fact, the density of the network is not a major factor, as routes are assumed to always exist and packets are assured a high delivery probability. Figure 4.6 shows that for a given frame distance, the MIR protocol outperforms the Back-to-Sink routing method by almost 1 hop for a faster mobile and 0.5 hops for a slower mobile. This difference arises from the extra hop required to return the packet the full distance back to the sink node. In networks covering a larger area, and possibly containing more nodes, these values will almost surely increase. This can be seen slightly in Figure 4.6, as a high network density is causing a larger gap between both routing methods as the mobile velocity is increased.

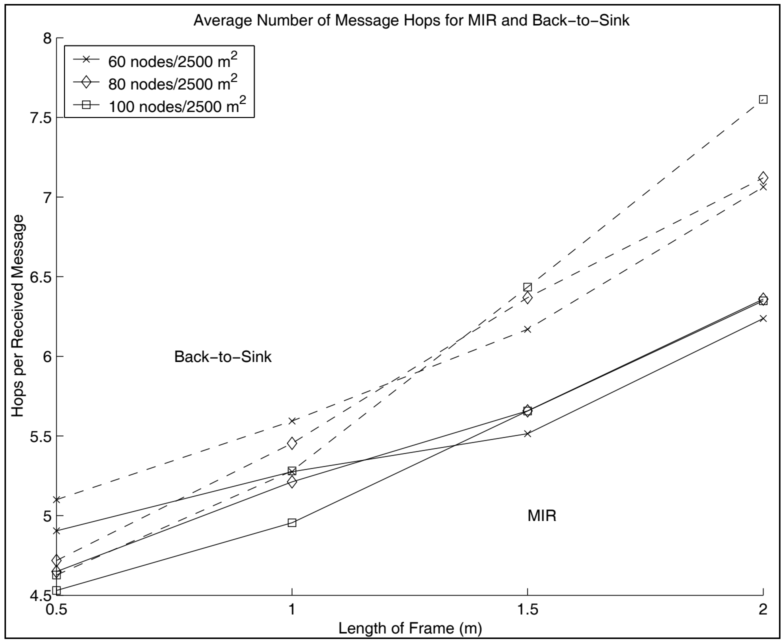


Figure 4. 5 Hop count for MIR vs. Back-to-Sink

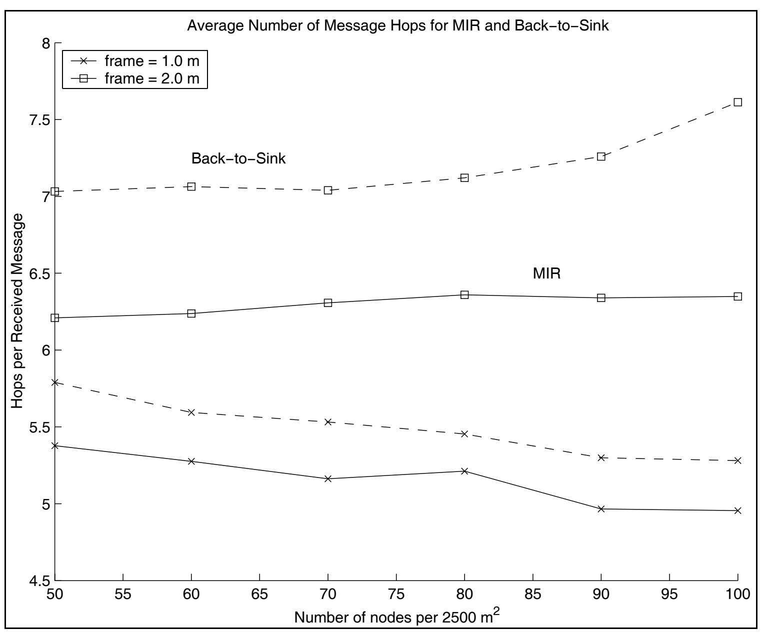


Figure 4. 6 Hop count for MIR vs. Back-to-Sink

4.4 Conclusion

A Hybrid Ad-Hoc Wireless Sensor Network (HANET) is presented, along with network level routing comparisons to cellular networks and MANETs, including hierarchical, proactive, and reactive techniques. Due to the presence of a stationary wireless backbone consisting of stationary nodes which are energy constrained, new network level routing protocols must be designed to provide packet routing support to and from mobile nodes.

The MIR protocol, or Mobile Intermediate Rerouting, is developed, which proposes to use the existing stationary sensor network and its multihop routing tree as a backbone to route message to the sink node from the mobile node. Essentially, as the mobile node may act as an information sink or a source, routing capabilities must allow for message delivery to and from mobile nodes. By using a control message initiated during the MAC level connection handshaking phase, a reverse route from the sink node to the mobile node can be formed. During further handoffs, only localized path updating is required, as specified in the MIR: Path Update algorithm. To avoid packet dropping, packet delay, signaling overhead, and message looping, an intelligent routing scheme is discussed in the MIR: Routing Algorithm.

The protocol, tested in PARSEC along side other MAC level protocols as well as in MATLAB. Results show that the MIR protocol is able to provide a low packet delay (signaling overhead) and high delivery probability as compared to other routing mechanisms in the face of mobility and varying network densities.

Chapter 5

Power Efficient Radio Control for Hybrid Ad-Hoc Networks Using Communication Outage Prediction

Here we consider ad-hoc networks. These are peer-to-peer networks which are comprised of tens to hundreds of similar nodes communicating through a wireless medium with ranges reaching hundreds of meters. The node topology of ad-hoc networks is either unknown at boot-up, or is continually changing. Protocol design for these networks usually depends on the assumption of similar mobility tendencies of the all of the nodes within the network, either stationary or mobile [3, 9, 15, 44, 52]. Hybrid Ad-Hoc Networks (HANETs) to be discussed herein allow for the more general interaction of stationary nodes and mobile nodes. On a functional level, these stationary wireless nodes can act as a backbone sensing network, providing multi-hop paths for data routed to a long range radio or sink node, while the mobile nodes can engage in tasks such as widespread environmental sampling or personnel security patrolling. For example, military personnel in defense applications can move through a sensor network extracting relevant information. Similarly, stationary sensors can be placed for area security with a small number of patrolling security officers. On a protocol design level, two different transparent protocols can be designed for each node type, allowing for different power saving techniques for mobile nodes.

The two cases of interest here are the interaction between mobile nodes and stationary nodes during Internal Connection Updating (ICU) and External Connection Updating (ECU). During the mobile node's dwell time within the communications range of the stationary node, the ICU functionality allows the two nodes to remain connected by updating the channel quality at the point of message reception. When no data is present, one node, usually with the responsibility of making connections, will send an ICU message, expecting a response with which to judge the channel. The ECU functionality allows the mobile node to selectively sense, or listen, to the channel for pilot signals (ECU messages) from nearby stationary nodes, the reception of which either involves entries into a neighbor registry or is a precursor to handshaking. Transmission of ICU packets (and ICU_Response messages) or reception of ECU packets involves energy usage. We propose an efficient radio usage scheme which allows the mobile radio to avoid unnecessary ICU and ECU functionality within a large class of MAC/Radio usage protocols, while continuing to provide a high connection probability.

Section 5.1 describes the general system model, including required MAC characteristics, radio channel model, and QoS prediction. Section 5.2 derives the efficient radio usage scheme in detail for the ICU and ECU problems. Section 5.3 provides simulation results and examples. Section 5.4 concludes the chapter.

5.1 System Model

5.1.1 MAC Characteristics

It is assumed that a wireless ad-hoc network has already formed and has reached steady-state. In many cases, the initial network formation is a set of stationary nodes. Either a specialized mobile MAC protocol will operate above the stationary MAC protocol or a single joint protocol will be running, defining the mobile-stationary interaction. The scheme presented here applies to many of these protocols, as it focuses only on the interaction of stationary and mobile nodes, and is not affected by the method by which the MAC level protocols handle carrier frequency selection, channel assignments (dynamic or static), power control (open or closed loop), channel contention, or collision detections. What is required is that the nodes are able to adjust their own schedule to turn their radios on and off. For convenience, we will assume that channels are formed by assigning a carrier frequency (or frequency hopping pattern) and a slot pair in a TDMA-type frame structure mutually agreed upon by the node-pair during a handshaking procedure. Without loss of generality, the frame length of the mobile node can be an integer fraction of the frame length of the stationary node, but we will assume it is the same length.

Some general characteristics are assumed of the MAC protocols governing a stationary and mobile node interaction. Before any connection can be formed between the two nodes, the stationary node will send out a pilot message, or a BROADCAST_INVITE, once every frame length on a common frequency. The mobile node can tune to this frequency at any time (according to the protocol) to listen for local stationary nodes and either register them as neighbors, or initiate handshaking. After a connection is formed, at least one slot pair (uplink and downlink) is reserved at each node for communications, with one node responsible for decisions to maintain the connection (Master). The other node is only responsible for transmitting data, responding to ICU messages, and listening for a disconnection message (Slave). Connection maintenance may be accomplished by determining the received signal quality based on received bit error rate. If the received signal quality drops too low, the Master node will transmit a disconnection message, usually at a higher power to accommodate disappearing Slave nodes. In the presence of data transfer, channel quality updating is performed on a regular basis. In instances of infrequent data transmission, the Master node will need to prompt the Slave node for an ICU message with the sole purpose of updating the link's quality.

5.1.2 Radio Model

We now describe our general radio channel model. Based on the received bit error rate, each node is able to estimate the quality of the link in terms of its signal to noise ratio (SNR). Also, we assume that the communications links will choose orthogonal channels, such as frequency or spreading codes, to mitigate interference from other links. Besides free space loss, we suppose there is environmental randomness, such as shadowing and multipath fading. This randomness can be encompassed by one random variable, Φ, for which a mean is given as $\overline{\Phi}$. Then, in general, the received SNR can be written as:

$$SNR_{rec} = \Phi \cdot \left(\frac{1}{k_0 \cdot B} \right) \cdot \left(\frac{r_{ref}{}^{n} \cdot G}{T} \right) \cdot P_t \cdot r^{-n} \tag{5.1}$$

Here, B is the signal bandwidth, P_t is the transmission power, r is the distance between the receiver and the transmitter, n is the free-space exponent, k_0 is Boltzmann's constant, and r_{ref}, G and T are constants associated with the antenna gain. This can be reduced to:

$$SNR_{rec} = \left[\frac{1}{B} \cdot r^{-n} \cdot P_t \cdot \Phi \right] \cdot \Psi \tag{5.2}$$

where Ψ encompasses all constants, antenna attributes, or factors whose control is out of the scope of this paper. Equation 5.2 can further be reduced to simply:

$$SNR_{rec} = \left(\frac{B'}{B} \right) \cdot \left(\frac{P_t}{P_t'} \right) \cdot \left(\frac{r'}{r} \right)^{n} \cdot \left(\frac{\Phi}{\overline{\Phi}} \right) \cdot SNR_{req} \tag{5.3}$$

This represents a scenario where the bandwidth, transmission power, and distance are allowed to vary. B', P_t' and r' are all reference terms which define a situation where the required SNR for reception, SNR_{req} , is attained when $\Phi = \overline{\Phi}$. If we assume that the bandwidth and transmission power are constant, the relation becomes:

$$SNR_{rec} = \left(\frac{r'}{r} \right)^{n} \cdot \left(\frac{\Phi}{\overline{\Phi}} \right) \cdot SNR_{req} \tag{5.4}$$

In this equation, r' represents the range of the transmitting node given a transmission power, bandwidth, and a required SNR. In a more general case involving power control methods, the transmission power at the stationary node will change, as well. This variation can be applied to the discussion here by using equation 5.3. We will assume, for simplicity, that the transmission power is held constant.

5.1.3 QoS Prediction and Estimation

In wireless networks involving mobile nodes, the main Quality of Service (QoS) measurements are outage probability, signal quality, and resource usage. QoS prediction techniques use one of these measurements, usually current received signal quality, to estimate future QoS values, attempting to improve overall system performance. The majority of prediction schemes in use today fall into one of three categories: location estimation and tracking [6, 19], cell coverage transition prediction [7, 30], and next step received power prediction for power control [56]. In each case, the signal quality of a received packet is used to predict possible future location or received signal quality and to allocate resources for future communications (transmission power, bandwidth) to prevent connection outage.

A novel variation of the idea of QoS prediction is to estimate the connection time from the received signal quality and reallocate resources used in maintaining connection information. For simplicity, cellular networks and sensor networks tend to assume that connection outages are due to exceeding the communication range of a transmitter. Thus, they employ feedback techniques to boost transmitter power [56] and soft-handoffs at the cell boundaries [55, 57]. Under these assumptions, given the transmission power, the received signal power, and the location/velocity information of the mobile node, it would be a simple matter to estimate the connection time, and forgo any extraneous signaling with the intent of maintaining connections until a cell boundary was reached. In [27], a method of determining connection time between two mobile nodes is investigated, with the goal of determining route longevity. It is assumed that both nodes have knowledge of position, velocity, and direction information. By adjusting the parameters to allow for one stationary node, this formulation breaks down to simply determining the distance to the cell boundary.

In practice, communication outages will occur within the transmission range of a cell due to environmental randomness caused by shadowing and multipath interference. Furthermore, it is unlikely that either node will have information regarding the mobile node's position and direction relative to the stationary node. (It can be assumed that the mobile node can determine its own velocity). Thus, determining the time until reaching the cell boundary becomes an unattractive solution to predicting outage. Instead, the idea proposed here is to directly predict the distance to the next connection outage by using the current received signal quality, and perhaps a time series of past signal qualities. Then, if it is determined that the connection will remain intact for multiple frame lengths with a certain confidence, the mobile node can conserve resources used in signaling to update, and search for, connections.

5.2 Efficient Radio Usage

5.2.1 Internal Connection Updating

The aim of any efficient connection scheme is to provide a temporary connection between two nodes, while limiting the signaling involved in forming, and maintaining these connections, thereby reducing the bandwidth and energy consumption. The decision to form connections is generally based on the received signal SNR at the mobile receiver, and the decision to maintain the connection is based on subsequent channel quality measurements made at the Master node during packet receptions. At the fringes of the communications range of the stationary node, bandwidth efficiency and the “ping-pong” effect avoidance are achieved by updating the channel frequently (at least once per frame) and setting a higher signal quality requirement for connections (to limit the effects of random environmental phenomena).

Once a connection is formed, however, frequent updates may not be necessary as the node pair may experience a strong connection and assume that the mobile will be in the vicinity of the stationary node for many frame lengths. The Master node retains a registry which contains information particular to each connection, such as node ID, last transmission time, connection status, and received SNR, which is continuously updated with each received packet. In general, either the mobile node or the stationary node can act as the master node, with the ICU functionality corresponding to one message transmission from both nodes. Thus, either configuration will result in energy usage at the mobile node, as well as the stationary node.

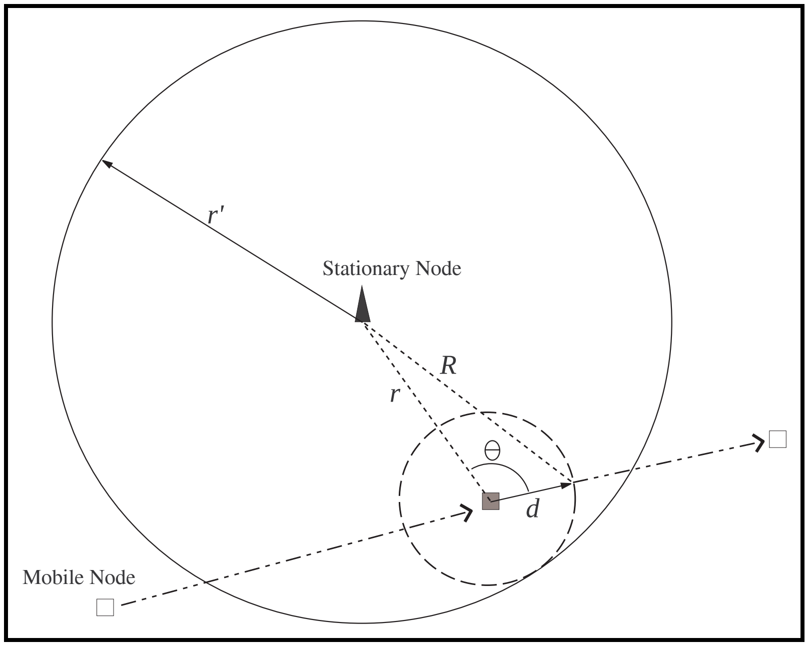


Figure 5. 1 Sample ICU functionality

We propose that once the mobile is connected to the stationary node, the Master node is able to predict with certain confidence that the mobile will remain within connection range even after moving some predicted distance. If the mobile node is the Master node, it will be able to make a prediction, within a confidence factor, of how far it can move from its position at message reception before it will be out of communications range, thus yielding an acceptable delay between ICU messages. Similarly, if the stationary node is the Master, given the mobile node's speed, it can predict the time it will take for the mobile to move the determined distance, thus giving the same acceptable delay. Figure 5.1 shows a possible predicted distance at the mobile node.

By examining the possible motion of the mobile node as it moves freely in relation to the stationary node in figure 5.1, we see that the future distance between the stationary and mobile node is given as:

$$R^2 = r^2 + d^2 - 2rd\cos\theta \tag{5.5}$$

Assuming the worst possible movement case, the mobile node will move away from the stationary node ($\theta = \pi$), yielding:

$$R = r + d \tag{5.6}$$

If the required received SNR for maintaining communications is known, then the probabilistic equation characterizing the ability of the node pair to maintain communications at a future separation distance is given as:

$$P\left(SNR\left(r+d\right) > SNR_{req} \mid SNR\left(r\right) = S\right) = \beta \tag{5.7}$$

Here, β is the confidence factor. A higher confidence factor will yield a more conservative movable distance. Using equation 5.4, we can reduce the above equation as follows:

$$\begin{aligned} S &= SNR\left(r\right) \\ &= \left(\frac{r'}{r}\right)^n \cdot \left(\frac{\Phi^-}{\overline{\Phi}}\right) \cdot SNR_{req} \end{aligned} \tag{5.8}$$

where Φ^- denotes the value of the random environmental variable at the first message reception. This easily reduces to,

$$r = \sqrt[n]{\frac{SNR_{req} \cdot r'^n}{S \cdot \overline{\Phi}} \cdot \Phi^-} \tag{5.9}$$

Using equation 5.8 to determine $SNR(r+d)$, and combining with equation 5.7,

$$P\left(SNR\left(r+d\right) > SNR_{req}\right) = P\left(r'^n \cdot \Phi^+ > \overline{\Phi} \cdot \left(r+d\right)^n\right) \tag{5.10}$$

By combining with equation 5.9, this simplifies to:

$$
\begin{aligned}
P\left(SNR(r+d) > SNR_{req}\right) &= P\left(\Phi^+ > \frac{\overline{\Phi}}{r'^n}\cdot\left(\sqrt[n]{\frac{SNR_{req}\cdot r'^n\cdot\Phi^-}{S\cdot\overline{\Phi}}}+d\right)^n\right) \\
&= P\left(\Phi^+ > \left(\sqrt[n]{\frac{SNR_{req}}{S}\cdot\Phi^-}+\frac{d}{r'}\cdot\sqrt[n]{\overline{\Phi}}\right)^n\right)
\end{aligned}
\tag{5.11}
$$

or,

$$
P\left(SNR(r+d) > SNR_{req}\right) = P\left(\Phi^+ > \left(\sqrt[n]{\gamma\cdot\Phi^-}+\Delta\sqrt[n]{\overline{\Phi}}\right)^n\right) \tag{5.12}
$$

$$
P\left(\Phi^+ > \left(\sqrt[n]{\gamma\cdot\Phi^-}+\Delta\sqrt[n]{\overline{\Phi}}\right)^n\right) = \beta \tag{5.13}
$$

where:

$$
\begin{aligned}
\gamma &= \frac{SNR_{req}}{S} \\
\Delta &= \frac{d}{r'}
\end{aligned}
\tag{5.14}
$$

are the relative SNR and the relative distance (r' is the maximum transmission distance of the stationary node). Φ^+, Φ^- are independent identically distributed random variables with mean $\overline{\Phi}$. Note here that no correlation is assumed between Φ^+ and Φ^-, as such information may not be available to the nodes.

Using this probabilistic equation and the environmental attributes such as fading distribution, required SNR, transmission distance, and free-space exponent, given a received SNR and a confidence factor, a corresponding movable distance can be found. If the velocity is assumed to be constant, a time until the allowable distance boundary can be obtained. Thus, the Master can avoid requesting an ICU message until the corresponding slot before the time expires.

5.2.2 External Connection Updating

A constant function of mobile nodes within a stationary network is to attempt to receive pilot signals from prospective stationary connection candidates. By maintaining connections to two or more nodes, the mobile node is protected from network outage due to a connection failure. Soft-handoff, for example, allows a mobile node to connect to a second node as it approaches the fringe of its first connection. Unless all of the stationary nodes are frame synchronized with fixed slots for pilot signals, the mobile nodes will need to continuously monitor the channel for possible connections.

In hybrid ad-hoc networks, it is required that the stationary nodes are able to form wireless links amongst themselves. Thus, the overlap of transmission ranges does not occur solely at the fringes of the coverage area, but will extend to at least the next stationary node. This may result in the possibility of having multiple connections which are not required. As the mobile node receives an ECU message, it will either add the stationary node as a neighbor (possible future connection) or as a connection (initiating handshaking). (It is assumed here that a stationary node's pilot signal is referred to as an ECU message until it is connected to the mobile node, after which ICU functionality takes over). A mobile node, therefore, may have an extremely high quality connection with one node, but will continuously monitor, and possibly connect to, other nodes in the network, resulting in energy usage due to channel monitoring.

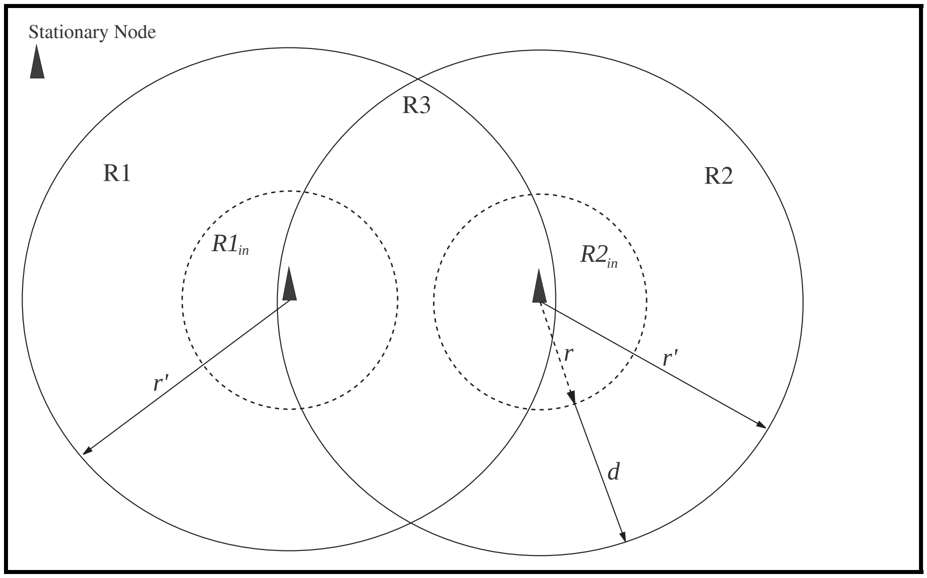


Figure 5. 2 Sample ECU functionality

We propose that an inner radius exists, well within the coverage border of a stationary node, within which a mobile node will not need to monitor the outside world for new connections. As the mobile node will only register updates of the received SNR of messages from the stationary node, the problem becomes that of the prediction accuracy of the time it will spend within this inner radius. When within the inner radius, the mobile node will forgo any monitoring for ECU message. As it exits the inner radius, while still within the transmission range of the stationary node, the mobile node will monitor the channel. Without loss of generality, the responsibility of received channel quality monitor can fall onto the stationary node or the mobile node. If the stationary node takes the responsibility, it will need to inform the mobile node of its monitoring requirements based on the received SNR.

Figure 5.2 gives a possible overlap of transmission radii of stationary nodes, with the proposed inner radii. In the absence of these inner radii, the mobile node would be able to connect to two nodes when located in region $R3$. But, in general, if the mobile has no knowledge of the boundaries of $R3$, it will continue to monitor for ECU messages while in regions $R1$ or $R2$. It is possible, though, to limit the area covered by $R1$ and $R2$ by adding the proposed inner regions $R1_{in}$ and $R2_{in}$. We then adjust the rules to prohibit ECU message monitoring while in regions $R1_{in}$ and $R2_{in}$.

Soft handoff methods are only possible when the mobile has the opportunity to connect to a second stationary node while maintaining at least one other connection. A brief network outage may be realized if this first connection is not maintained when an ECU message is received. Thus, it is important for the mobile to remain in the overlap region long enough to be able to connect to both nodes, usually the amount of time required to handle one full handshaking procedure. We employ a similar strategy in our case. The mobile node must be required to spend enough time in the outer regions ($R1$ or $R2$) to allow the MAC protocol to obtain a full sense of the nearby network, and possibly form new connections. In the worst case scenario, this would include one full frame to receive an ECU message, and another frame to establish a connection. Effectively, if the mobile spends these two frames outside of the inner radii, it will revert back to a state where it never stopped sensing.

This becomes a problem of determining a distance from the transmission border which, given the velocity of the mobile node, allows it to remain in the outer ring for at least two frame lengths. Using the labels given in figure 5.2, the probability equation can be written as:

$$P\left(SNR\left(r+d\right) > SNR_{req} \mid SNR\left(r\right) = SNR_{in}\right) = \beta \tag{5.15}$$

where β is the confidence factor. If we assume that the frame length is given as t_f (seconds), and the velocity of the mobile is given as v (meters/second), then the equation can be rewritten as:

$$P\left(SNR\left(r+2t_f v\right) > SNR_{req} \mid SNR\left(r\right) = SNR_{in}\right) = \beta \tag{5.16}$$

Although this equation looks similar to the ICU derivation, it differs in that the value of SNR_{in} is unknown, and must be determined with the knowledge of t_f and v. In the ICU case, the value of SNR_{rec} was used to determine d. By comparing this to equation 5.7, a simplification can be written as:

$$P\left(\Phi^+ > \left(\sqrt[n]{\gamma \cdot \Phi^-} + \Delta\sqrt[n]{\overline{\Phi}}\right)^n\right) = \beta \tag{5.17}$$

where

$$\begin{aligned} \gamma &= \frac{SNR_{req}}{SNR_{in}} \\ \Delta &= \frac{2t_f v}{r'} \end{aligned} \tag{5.18}$$

Using this equation, in the case of no randomness in the environment, $\Phi^+ = \Phi^- = \overline{\Phi}$. Thus, the relationship between SNR_{in}, t_f, and v can be found as:

$$SNR_{in} > \frac{SNR_{req} \cdot r'}{\left(r' - 2t_f v\right)^n} \tag{5.19}$$

suggesting that as the velocity or the frame length increase, the inner radius will decrease, thereby enlarging the outer ring within which the mobile node will need to begin sensing the world.

With knowledge of the required SNR, transmission distance, free-space exponent, fading distribution, and mobile node attributes such as velocity and frame length, an inner radius can be found within the transmission range of the stationary node. Within this inner radius, the mobile can confidently avoid sensing the outside world for ECU messages, thereby conserving power associated with radio level receptions.

5.3 Results

We are interested in characterizing the relationship between γ, Δ, and β, as well as their behavior under the changes in frame length and mobile node velocity. As the solution for probabilistic functions are not generally trivial, simulations are used to generate plots, and derive numerical results.

As the derivation for the power control schemes in the ICU and ECU cases was remarkably similar, it is not surprising that the effects related to varying user definable parameters is also similar. In the environments we are interested in, we assume that the mobile node's velocity and frame length are parameters which are controllable. All other quantities, such as the environmental random distribution, the required SNR, and the free-space exponent are assumed to be constant. For the ICU case, as the value of either t_f or v are decreased, more ICU messages will need to be passed within a fixed time duration or distance, respectively. Thus, by using the power saving scheme and suspending ICU messages for a distance d, a higher power saving is attained. For the ECU case, shorter frame lengths and lower velocities result in a lower value of the relative distance Δ. As Δ and γ are inverse related, this results in a decrease of SNR_{in}, and a larger area covered by the inner radius of the stationary node. Thus, the ECU message receptions are suspended for distances further away from the stationary node.

Numerical simulations were performed depicting the relationship between γ and Δ. The environmental random variable is assumed to be a lognormal random variable (underlying Gaussian variable with a standard deviation of 8 dB). We assume that the randomness caused by multipath interference can be mitigated using diversity and error control techniques [53]. Figure 5.3 gives a realization for various values of β for $n = 4$. Note that when plotted on a logarithmic scale, γ is almost linearly related to Δ with a negative slope which remains constant as β varies. This is due to the approximation of the logarithm of γ to the value of $\sqrt[n]{\gamma}$ when γ is small. Also, note that as β decreases, denoting a lower confidence level, and either Δ or γ is held constant, the other will increase. In the ICU case, where Δ must be determined, this results in an increase in the allowed movable distance before connection updating. The mobile will be allowed to conserve energy for a longer duration, but may lose its connection at reaffirmation. In the ECU case, an increase in γ will lower the value of SNR_{in}, enlarging the area of the inner region. Again, the mobile is allowed to forgo ECU monitoring, but will suffer the possibility of moving through the outer ring too quickly.

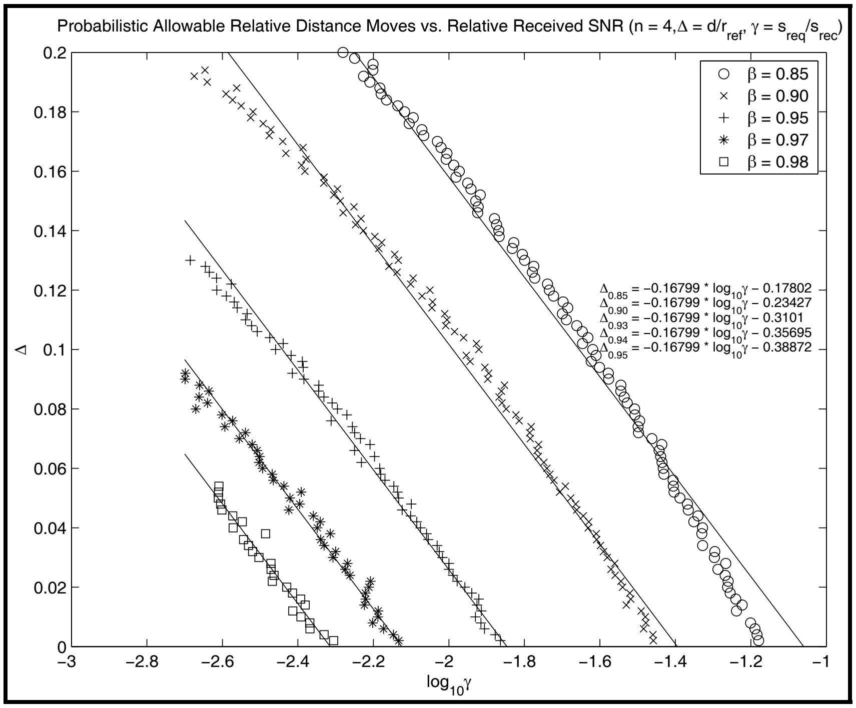


Figure 5. 3 Beta curves for radio control

For the analysis of ICU and ECU message transmission, the performance was measured by investigating the ratio of the number of messages transmitted using the radio control scheme versus the number of messages transmitted under normal conditions. Assuming that under normal conditions, the mobile will transmit one ICU message and receive one ECU message per frame, the ratio is simply the average number of either message transmitted per connected frame. Also, the shadowing randomness is assumed to be autocorrelated with a 5 meter reference distance according to the model presented in [17].

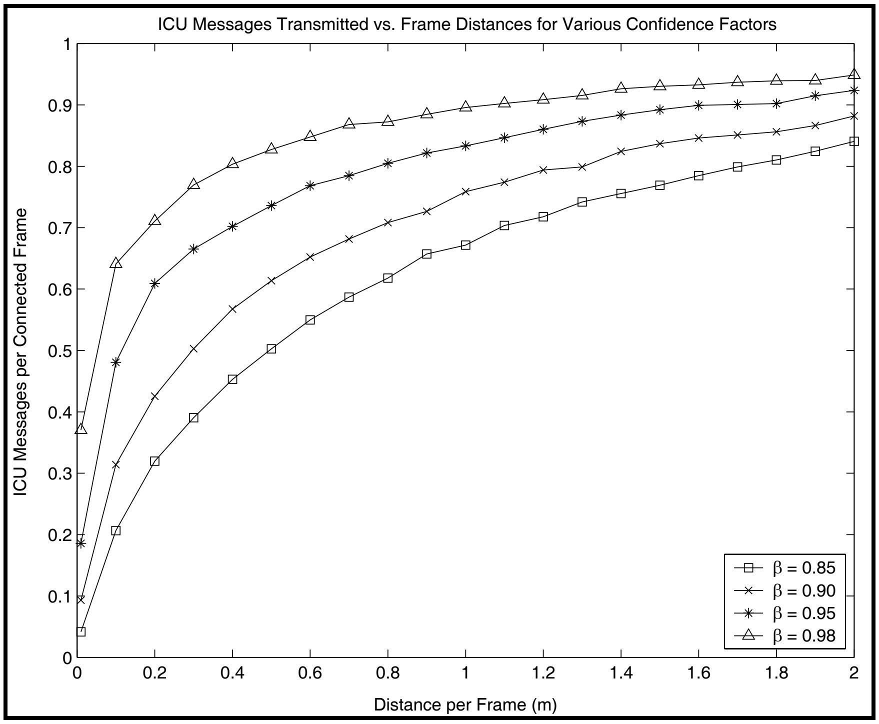


Figure 5. 4 ICU signaling rate

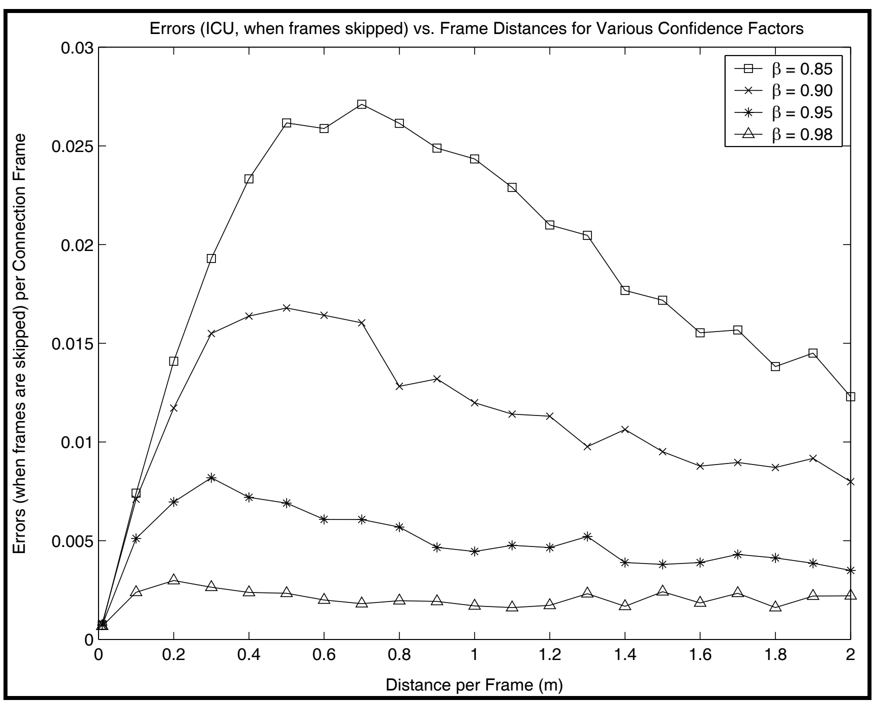


Figure 5. 5 ICU errors per frame

Figures 5.4 and 5.5 show the results for the transmission of ICU messages for various confidence factors. To overcome harsh shadow fading, a fade margin of 10 dB was used. Also, we assume binary-PAM transmission requiring a bit error rate of 10^{-3}, yielding a required SNR of 4.77. To simulate signaling requirements, mobile nodes were introduced outside of the transmission range of a stationary node, and were allowed to move towards the stationary node, form a connection, and finally disconnect. At each connected frame, a decision was made as to the next update frame. Errors occurred as the mobile node incorrectly predicted a time to update a connection, and entered an outage region. As expected, the number of ICU messages transmitted per frame approaches the regular frequency of once per frame as the confidence factor approaches 100 %. Also, for faster moving mobiles, or mobiles with longer frame lengths, a higher frequency of ICU messages is required. Figure 5.5 shows the error rate (per connected frame) when frames are skipped. Therefore, when no frames are skipped, it is assumed that the error was not due to negligence caused by forgoing updates.

Figures 5.6 and 5.7 show the results for the reception of ECU messages when an inner region is used to avoid external monitoring. The algorithm here was compared to a standard soft-handoff algorithm. A mobile node initiates a “call” within the range of one stationary node, and proceeds into the range of the next stationary node. We assume a distance between stationary nodes equal to its own reference range, allowing wireless communication between both stationary nodes as well. The outage probability of the connection under the radio control scheme is compared to that of the standard soft-handoff scheme which continuously monitors for pilot messages. Again, if we assume that the mobile will monitor for pilots once per connected frame, the ratio will denote the average number of receptions per connected frame. Figure 5.6 shows the average number of receptions per connected frame. Figure 5.7 depicts the factor by which the outage probability is increased. Note that even at low distances moved per frame time where the inner region is at its largest, the increase in outage probability is less than 5% with even a low confidence factor.

Figures 5.8 and 5.9 incorporate the idea of Kalman filter prediction of received signal power, presented in [48], to the ICU problem. Here, the signal quality at the next measurement instant (before it is determined to skip the frame) is estimated using a Kalman filter, taking into account variations in lognormal shadowing (with driving noise variance of $2.0\,dB^2$ and local mean measurement errors variance of $9.0\,dB^2$). Similarly, a simple unweighted linear predictor is used over the last 10 samples. (Note that even though the Kalman filter assumes no skipped measurements, we can still use it in the ICU case when frames are skipped with a small cost of prediction accuracy). The Kalman filter estimate, which provides a conservative estimate, is compared to the linear prediction estimate. In the cases where the linear predictor yielded a higher estimate than the Kalman filter, the mobile node was assumed to moving towards the stationary node, and the mobile was allowed to move twice its distance before updating. Otherwise, it is assumed that the mobile is moving away from the stationary node, and the algorithm proceeds normally. There is an improvement in ICU message transmission rate at the expense of error rate per connected frame. This analysis suggests that as the error rate using direction prediction is still low (under 1% for the mobility patterns considered here), there is a clear benefit to using a past time series to reduce the power associated with the transmission of ICU messages.

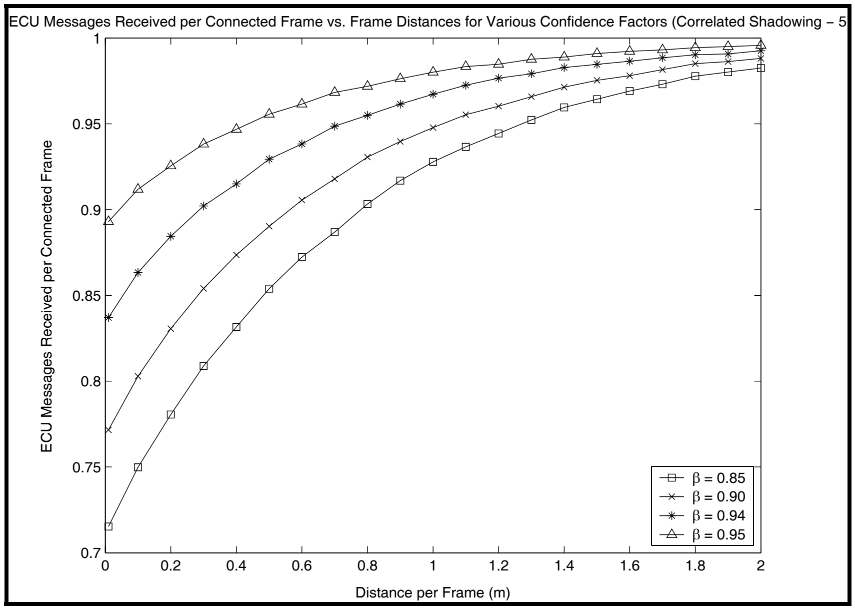


Figure 5. 6 ECU signaling rate

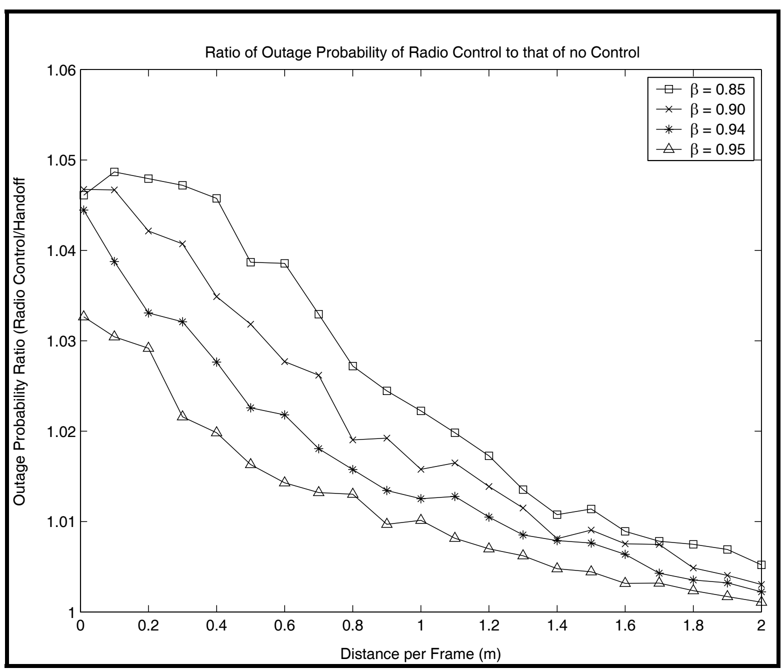


Figure 5. 7 ECU outage probability

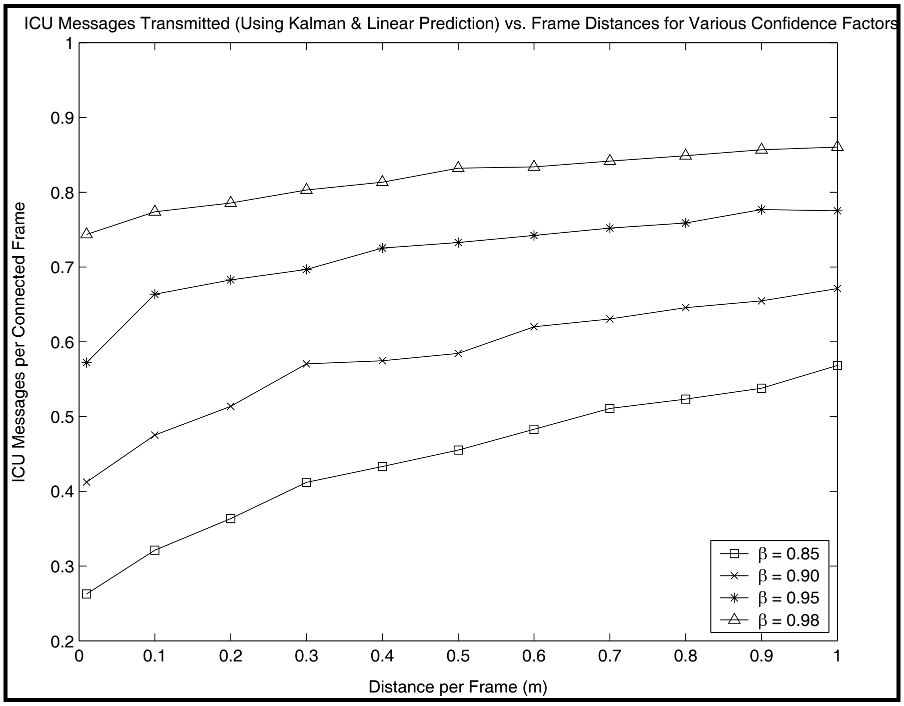


Figure 5. 8 ICU signaling rate (using Kalman prediction)

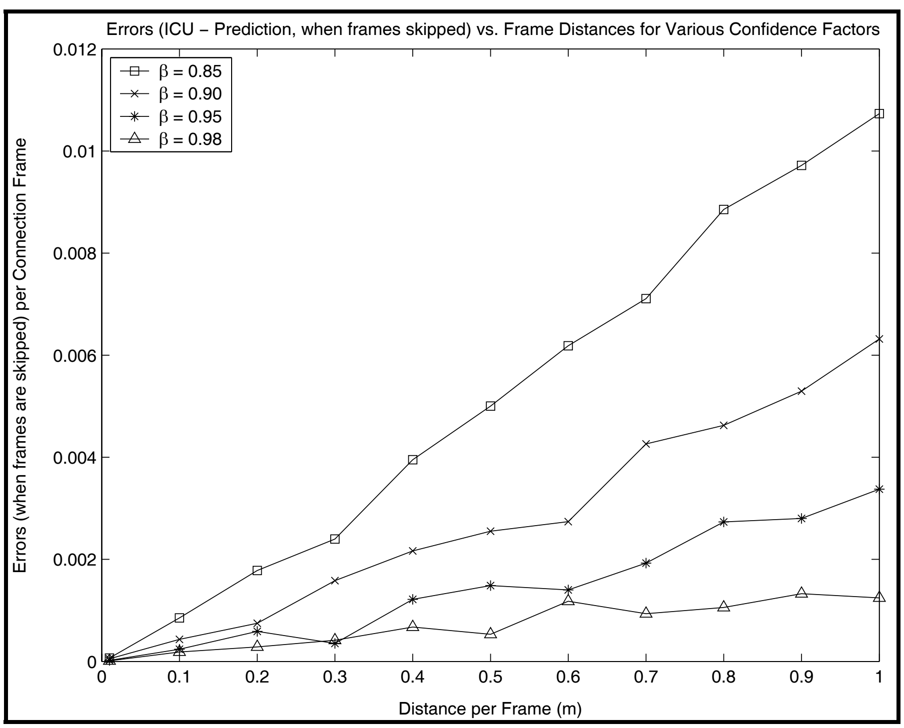


Figure 5. 9 ICU errors per frame (using Kalman Prediction)

5.4 Conclusion

A new concept of a hybrid ad-hoc network is presented. We illustrate two message exchanging functions which employ ECU and ICU messages to establish and maintain connections between stationary and mobile nodes. By avoiding repeated connection updates and unnecessary pilot message channel monitoring, it is possible to reduce the amount of energy used due to receiving and transmitting these messages.

By exploiting the expected distance to the transmission fringe of a connected node, a method to delay connection updates is derived as a probabilistic equation, depending on received SNR, confidence factor, and certain environmental attributes. The result does not depend on MAC level protocol specifics, and the results scale with transmission distance and required SNR. The confidence factor will govern how conservative the movements will be. Then, by exploiting the proximity of the mobile node to the stationary node, the probabilistic equation derived for the suspension of pilot channel monitoring was shown to be similar to the earlier result.

We conclude that for scenarios where multiple connections are possible, these power control schemes are able to avoid connection maintenance and formation whenever possible if reliable connections are present. The benefits of using such schemes become more apparent as the frame length or the mobile node velocity is reduced. Furthermore, as past signal measurement information is used, there is an improvement in energy efficiency at the slight cost of increased error rate.

Chapter 6

D-PEC: an Energy Efficient Distributed Pre-Event Clustering Algorithm

Wireless sensor networks consist of a dense, possibly randomly distributed, group of sensors interconnected via a wireless medium to perform sensing tasks. Prior to deployment, it is assumed that nodes have no knowledge of relative location, channelization, or routing capabilities. It is possible, though, that nodes within this ad hoc network have control of their own radios. Before sensing operations can begin, the nodes will form a MAC level link architecture and a multihop routing scheme to propagate messages to a sink node or long-range radio. The nodes themselves are considered to be homogeneous, with no node assumed to have higher resources or processing power associated with it. The lack of a wired infrastructure and the ability of the nodes to boot up and form network level routes [14, 51, 52] suggests a wide range of applications where human interaction may not be possible due to inhospitable or inaccessible terrain. This type of network is envisioned to have applications such as military defense, target localization, and environmental sampling. Without the possibility of interaction, though, it is inevitable that the nodes will be depleted of energy supplies. Thus, the primary goal of ad hoc wireless sensor networks is to prolong network lifetime by conserving energy at each sensor node.

In addition to costs associated with sensing tasks, energy consumption occurs in two main areas: data processing and radio transmissions. As described in [43], in wireless networks, communications is the major consumer of energy. As an example, it costs 3 J of energy to transmit 1 kb of data a distance of 100 m. A general purpose processor, on the other hand, operating at 100 million instructions per second (MIPS)/W executes 300 million instructions for the same amount of energy. Thus, to preserve energy at the node level, it will be necessary to allow for localized processing of information as opposed to reporting raw data to the sink node. In densely populated sensor networks, tasks such as target identification will involve the participation of multiple nodes, and thus some communication of raw data is unavoidable. Allowing each node to report raw or processed data to the sink node via multi-hop channels will (a) cause an energy drain due to network traffic and (b) create a bottleneck of information at the nodes surrounding the sink node. As an alternative, it has been proposed that stationary clusters can be used to process data from multiple nodes at the local level, increasing the validity of sensor measurements and reducing the traffic overhead associated with reports.

Many algorithms have been investigated to efficiently combine data at a clusterhead node with the purpose of data fusion or beamforming [5, 6, 19]. Traditionally, the targets of interest have been mobile in nature, e.g. vehicles or personnel. Upon sensing a target, multiple clusters will either identify the target, determine its bearing from the cluster, or possibly both. To provide meaningful and timely data, the latency between event occurrence and target/location identification, including cluster formation, local data routing, and data processing, should be low.

Furthermore, as mobile target events tend to spatially correlated, certain nodes will be involved with clustering repeatedly, suggesting a need to reduce the signaling overhead associated with cluster formation and data routing.

Data processing algorithms, while assuming the sensor nodes know their own location, do not investigate methods of cluster formation. Until recently, it has been assumed that sensors can be manually configured and placed in optimal positions to aid in data processing tasks. This is not the case for ad hoc wireless sensor networks. As node location is generally assumed to be random, as is the number and location of node neighbors, forming clusters prior to deployment is nearly impossible. Furthermore, centralized formation of clusters after deployment will involve each node to report its location and link status via the multihop network. This becomes infeasible as the network size grows to 100's and 1000's of nodes. We propose a clustering scheme which allows a network to form clusters in a distributed fashion with each node having knowledge of its own link level connection information. Furthermore, our scheme will be shown to be effective in conserving energy supplies when used to identify events which may be mobile or frequently occurring by eliminating the overhead associated with event-driven cluster formation and reducing the overhead associated with local cluster routing.

Section 6.1 describes clustering design issues and current clustering schemes. Section 6.2 develops the pre-event based clustering scheme. Section 6.3 gives analytical bounds for the signaling overhead associated with our clustering scheme, as well as comparisons to current competitive schemes, with section 6.4 concluding the chapter.

6.1 Clustering Design Issues and Current Strategies

Various considerations must be addressed prior to the inclusion of clustering into any networking architecture. These fall into three classes:

Power Requirements: In either event-driven (temporary clusters are formed as events occur) or pre-event driven (permanent clusters are formed prior to events) clustering, there are three stages of cluster usage, each with power requirements which involve a depletion of energy at all nodes within the cluster. Formation of a cluster, during which time a clusterhead is elected and members are assigned, involves some degree of signaling and time delay. After an event is detected, all member nodes will route data to the clusterhead, possibly via multihop paths if the cluster topology is complicated. Finally, the clusterhead node will incur the cost of data processing.

Network Attributes: In deciding which nodes are to be clusterheads, each node's capabilities must be identified. Nodes with a higher energy reserve, for instance, are more suited for processing data. Alternatively, it is possible that all nodes are homogeneous. In such cases, it is not clear as to which nodes are to be elected as clusterheads. If the nodes are randomly distributed with no network knowledge prior to deployment, there must be some method by which they can extract the required metric which is the basis for the clustering scheme, such as link-level connectivity and neighbor signal quality.

Clustering Goal: As network architectures vary from fully mobile (MANETs), fully stationary (WINS), to possibly an interaction between stationary and mobile nodes (HANETs), it is important to identify the goals of the clustering scheme. In MANETs, for instance, clustering is used to facilitate network organization and routing requirements [2, 15]. Clusterheads are elected to provide a hierarchical backbone in connecting nodes from one cluster to another. In stationary networks, as those investigated in this study, clusters are formed to provide more effective sensing abilities by clustering data. Furthermore, if it is the case that nodes are energy limited, clustering can provide reduced network traffic, and a longer network lifetime.

The rest of this section will focus on various schemes which allow for stationary network clustering for the purposes of data processing.

6.1.1 Centralized Methods

In situations where physical placement of nodes is possible, optimization of the clustering architecture is performed in a centralized fashion. Locations of the nodes are pre-planned, and the issues regarding MAC level channelization and network level routing are solved prior to node placement. Also, as the clusterhead has already been selected, it can be given a more powerful processor along with higher energy reserves, allowing for possible long range single-hop communications to a sink node. Unfortunately, as the assumption suggests, this method requires user placement as well as prior knowledge of location of the nodes.

6.1.2 Distributed Methods

Until recently, distributed methods to cluster stationary networks have not been shown much interest. With the supposition of a randomly distributed field of sensors possibly consisting of many hundreds of nodes, the abilities of centralized methods begin to break down, suggesting the need for more distributed techniques. The two types of schemes which have been studied recently fall into two groups: (a) those which use clustering as a substitute for multi-hop routing schemes and (b) those which use multi-hop event driven clustering schemes.

As an alternative to multi-hop communications, clustering can be used to form a hierarchical backbone of clusterheads for routing purposes. This is analogous to the clustering techniques of MANETs. In effect, the area encompassed by the nodes is segmented into smaller areas, each containing a clusterhead, the choice of which may be random. All nodes within any given area, then, belong to the same cluster, and report to the clusterhead in a single-hop, regardless of distance. The LEACH algorithm [18] proposes this method. In the LEACH algorithm, clusterheads nominate themselves randomly, and broadcast their intent to the entire network via a CSMA protocol requiring all other nodes to continuously monitor the channel to determine their optimal membership location. These clusterheads are then required to act as a one-hop bridge between the cluster members and the sink node. To avoid the burden on these clusterheads (which now act as gateways), the network will periodically re-elect clusterheads and determine new membership status for all nodes. Unfortunately, this reconfiguration imposes an energy drain on all nodes due to signaling, as well as a time delay due to cluster set up. Also, the locations and number of clusterheads cannot be guaranteed for any given network formation, suggesting a centralized method to efficiently span the network area. Figure 6.1(a) gives a realization of the LEACH algorithm.

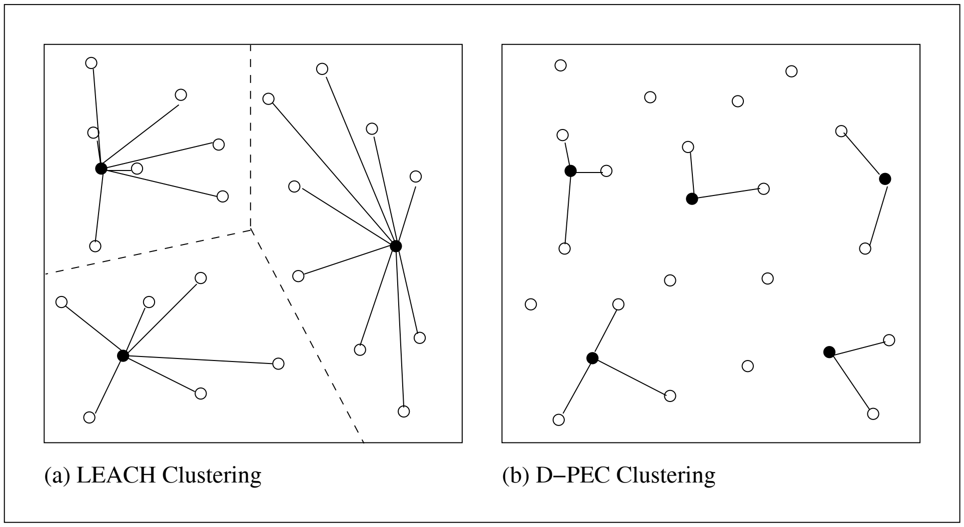


Figure 6. 1 LEACH vs. D-PEC clustering

In multi-hop event driven clustering, multi-hop routes are assumed to already exist linking any stationary node to the sink node. As events occur requiring multiple node cooperation, temporary clusters can be formed which allow data sharing, after which the nodes will assume normal operation. Single Winner Election (SWE) [14] adopts this technique. Nodes wishing to participate in cluster formation will declare intent to all neighbors. Short messages are then exchanged by all participating nodes to determine the most likely candidate to be elected as clusterhead, using metrics such as energy reserves or time of declaration. Data is then routed to the clusterhead, possibly via local multi-hop as the cluster configuration is not necessarily a simple star formation. For tracking of mobiles, though, it is possible that similar clusters will be formed continuously, incurring the energy associated with signaling on a repeated basis. Also, there is a time delay in forming the clusters, as an election process must take place and local routes need to be formed. Figure 6.2(a) gives a realization of the SWE algorithm.

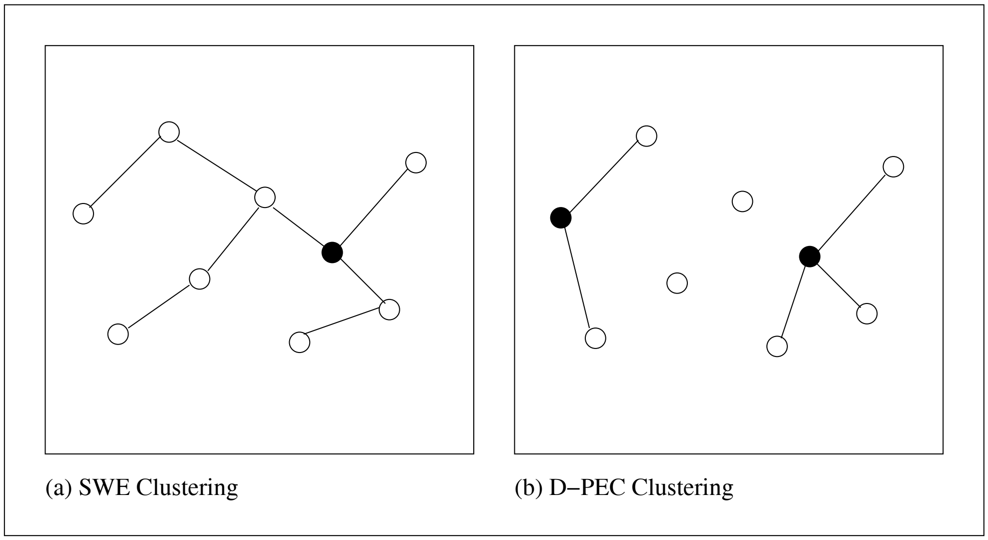


Figure 6. 2 SWE vs. D-PEC clustering

6.2 Distributed Pre-Event Clustering (D-PEC)

In ad hoc stationary wireless sensor networks, the primary resource is energy. Thus, in developing a clustering scheme to track mobile or frequently occurring targets, it is necessary to avoid energy usage due to signaling and long distance communications. We assume that the primary mode of routing information through the network is via multi-hop routes. Also, we assume that prior to any clustering scheme, the nodes have established a MAC level channel usage scheme, and links to neighboring nodes have been formed. To avoid the signaling associated with repeatedly setting up clusters or reconfiguring to rotate a small number of clusterheads, we adopt a pre-event based scheme which contains many clusterheads. We present here Distributed Pre-Event Clustering (D-PEC).

Immediately following the MAC level boot up phase, the nodes will enter the clustering phase. During the boot up phase, each node establishes a TDMA-type frame schedule. Slots are chosen from this schedule, as well as a frequency or frequency hopping pattern, to communicate with neighboring nodes. The choice of channels and the techniques by which link-to-link interference is mitigated is irrelevant, and out of the scope of this study. The goal is to allow the energy constrained network, in a distributed fashion, to set up small clusters, of user definable sizes, which are evenly distributed throughout the network area. The metrics chosen for clusterhead determination are degree of connectivity, studied in [36], and lowest node ID [12]. Prior to deployment, each node is instructed as to the maximum and minimum number of members per cluster (MAX_CLUSTER_SIZE, MIN_CLUSTER_SIZE). The clustering algorithm proceeds through three phases: Cluster_Start, Cluster_Declare, and Cluster_Set. Messages transmitted within each phase are not broadcast throughout the network, as is in the LEACH protocol, but transmitted only to the receiving node via its predetermined channel. The protocol proceeds as follows:

Cluster_Start: Each node sets up an array of clustering candidates which includes all of its neighbors (with the size of the array as MAX_CLUSTER_SIZE – 1). If the number of neighbors exceeds this allotted size, a choice is made based on received signal quality (SNR). If the candidate list does not meet the minimum size requirement (MIN_CLUSTER_SIZE – 1), the node drops out of this round. For all nodes that reach qualifying status, a message is sent to all entries in the candidate list containing connectivity information (CLUSTER_INVITE). At this time, each node will have information as to the connectivity of all nodes within its immediate proximity, as well as its own. A winner is chosen from this group (possibly itself), with equal connectivity status contention being eliminated with lowest ID requirements.

Cluster_Declare: If a node has elected itself as a winner, it sends out a CLUSTERHEAD message to all nodes on the candidate lists, informing them of the winning node. Nodes receiving a clusterhead message compare the sending node with its own assumed winner. If the receiving node agrees with the clusterhead choice, a CLUSTER_ACK message is sent as a response to the winning node. The winning node accepts the response, and keeps a record of the reporting nodes. All nodes are given a chance to respond, following which each clusterhead node knows the number of accepting nodes from its candidate list.

Cluster_Set: For each winner (clusterhead) in the network, if the number of accepting nodes is greater than MIN_CLUSTER_SIZE – 1, then it has formed a successful cluster, and needs to inform the members that they may participate in cluster operations. Again, a CLUSTERHEAD message is transmitted. All nodes receiving this message again will register their group ID (the node ID of the clusterhead) as well as the outbound link to the clusterhead node.

Upon completion of the D-PEC algorithm, not all nodes in the network will be part of a cluster. But, the following three properties arise:

(1) Each Clusterhead will be the most connected node of the group.

(2) Each cluster will be in a "star" formation, with the clusterhead being the center node, and all members at a 1-hop distance.

(3) Each cluster will contain anywhere between the minimum and maximum number of members designated at deployment.

By properties (1) and (2), it can be seen that no winner is within the same cluster as another winner. Also, by all three properties, if any area in the network contains a group of nodes numbering between the required minimum and maximum count, the most connected node will almost always become the clusterhead (called a likely cluster). This suggests that the clusterhead nodes will be evenly distributed over the area of the network. Figures 6.1(b) and 6.2(b) give comparisons of sample final realizations of the D-PEC algorithm as compared to the LEACH and SWE algorithms.

Situations may arise, though, in which a likely cluster will not form. In one case, it is possible for a winning node (possible clusterhead) to have a node in its cluster candidate list which itself is the winner of another group (and hence, did not respond in round 2). In this case, both winning nodes may invite each other, but neither will respond with an acknowledgement message. Also, in another case, it is possible that Node A is not in the candidate list of Node B (due to size/invitation limitations). If Node A is isolated (its only connection to the network is via Node B), then it may not join Node B's cluster. Obviously, there is a distinct tradeoff between complexity and optimization. To reduce complexity, these cases are not resolved to an optimal clustering solution as there would be an increase in energy usage due to signaling during the resolution, which may not be required if the performance of the clusters is not significantly decreased.

6.3 D-PEC Signaling Overhead

As the resource of interest here is energy usage at the radio level, we propose that the main performance measurement of clustering algorithms in ad hoc wireless networks is energy depletion due to signaling. In networks where the link level architecture has been designed such that nodes communicate via orthogonal channels at fixed time intervals (using short hop communications), radio energy usage determination becomes a simple matter of enumerating the number of messages required to cluster the network. In link level architectures where information is broadcast throughout the channel via multiple access techniques such as CSMA, and via long distance transmissions, a method of comparison must be established.

6.3.1 Radio Level Energy Usage

For short packet messaging, it can be assumed that all packets are of the same length, requiring the same bandwidth, T (sec) and B (bits/sec). Thus, there will be TB bits per slot (or bits per message, if we assume one message per slot). When powered down, each radio is assumed to not dissipate any energy. When powered on, the dissipation energy is given as E_{diss} (J/bit). During transmission, the energy associated with signal amplification is E_{amp} (J/bit/m^4). This, coupled with a desired distance of transmission (r) gives the amount of additional energy required to transmit one bit a distance r in an environment dominated by a 4th power falloff in signal strength. Given the number of bits per message, the transmission and reception energies are:

$$
\begin{aligned}
E_t &= B \cdot T \cdot E_{diss} + B \cdot T \cdot r^4 \cdot E_{amp} \\
E_r &= B \cdot T \cdot E_{diss}
\end{aligned}
\tag{6.1}
$$

At short distances, on the order of 10 meters, and assuming probable radio specifications such as E_{diss} = 50nJ/bit and E_{amp} = 13pJ/bit/m^4, the transmission and reception energies are roughly equal, suggesting that most short distance, short length, transmissions will require roughly the same amount of energy. To simplify matters pertaining to algorithm comparison, we assign the symbols of $\widehat{E}_t$ and $\widehat{E}_r$ to short distance transmissions and receptions where:

$$
\widehat{E}_t \approx \widehat{E}_r \tag{6.2}
$$

In situations where a receiver will remain powered on longer than one slot, we will proceed to determine the number of slots required to remain on to receive a given message, and determine an equivalent number of single slot receptions which have taken place. On occasion, though, an algorithm may require a transmitter to communicate over longer distances than those dictated by the "short-distance" criteria. In such cases, the following definition can be used (see appendix B):

$$
\begin{aligned}
\gamma &= \frac{\Delta + \rho^{\alpha}}{\Delta + 1} \\
\Delta &= \frac{E_{diss}}{E_t^{'} \cdot fm} \\
\rho &= \frac{r}{r^{'}}
\end{aligned}
\tag{6.3}
$$

where γ is the number of equivalent short-distance messages that can be transmitted using the same amount of energy to transmit one message to a distance of r in an environment where α is the exponent for signal strength falloff. The value of $E_t^{'}$ represents the amount of energy required to transmit a message, including energy dissipation, to a distance of $r^{'}$. In a generalized shadowing environment, it may be necessary to include a fade margin, which is given as fm. Thus, given a reference short distance and relevant radio specifications, the value of γ can be determined which reduces a long distance transmission to a number of short distance transmissions.

6.3.2 Network Density Issues

At extremely short ranges, the primary cause of energy usage during communications is energy dissipation at the radio. Thus, it can be conceived that as the density of the network increase, and distances between nodes decrease, that the energy used in setting up clusters becomes immense unless bounds are set as to the allowable sizes of clusters. As the network density increases, therefore, the bounds on the maximum cluster size limit the number of invitation messages and node participation, and thus bounding the number of messages required per node to set up a clustered architecture. On the contrary, as the network density decreases, the number of connections for any given node will also decrease. The number of nodes participating in any given clusters' formation will then depend on the number of connections, and therefore the density of the network.

Utilizing the law of large numbers, we can estimate the number of connections for each node by the network node density, and each nodes average coverage area. Assuming the same notation convention used above for short distance transmissions, the coverage area of any given node is:

$$r_{\text{cov}} = E\left[r' \sqrt[\alpha]{fm} \sqrt[\alpha]{\frac{\Phi}{\overline{\Phi}}} \right] \tag{6.4}$$

where Φ is the environmental random variable, encompassing all of the shadowing and multipath interference, with $\overline{\Phi} = E[\Phi]$. Also, we note that:

$$\begin{aligned} E\left[\sqrt[\alpha]{\Phi}\right] &\neq \sqrt[\alpha]{E\left[\Phi\right]} \\ r_{\text{cov}} &= r' \sqrt[\alpha]{\frac{fm}{E\left[\Phi\right]}} \cdot E\left[\sqrt[\alpha]{\Phi}\right] \end{aligned} \tag{6.5}$$

Assuming the environmental random variable consists only of lognormal shadowing (with 8dB mean for the underlying Gaussian random variable), we determine that $E\left[\sqrt[4]{\Phi}\right] = 1.1129$ and $\sqrt[4]{E\left[\Phi\right]} = 1.5283$, giving a coverage radius of $r_{\text{cov}} \simeq 13$ when the fade margin is 10 dB. Multipath effects can be eliminated by assume high-order diversity [53]. Now, given a network area (A_{net}) and number of node (n), the number of neighboring nodes within a given node's coverage area is given as:

$$n_{\text{cov}} = \frac{n}{A_{net}} \pi r_{\text{cov}}^2 \tag{6.6}$$

We define the high density case, then, as:

$$n_{\text{cov}} > MAX\_CLUSTER\_SIZE - 1 \tag{6.7}$$

as the maximum cluster size that will include the central node as well as all of its neighbors.

6.3.3 Message Counting in D-PEC

In D-PEC clustering, we assume that nodes communicate via orthogonal channels separated in time and frequency. Also, we assume that, due to node proximity, any message that is transmitted will be received with no errors. Therefore, to determine the energy used in forming a clustered network, the number of total messages can be counted, and converted to equivalent energy usage statistics by using the relationships derived above.

As mentioned earlier, four message types are used to cluster the network during the three phases. During the first phase, each node sends out a CLUSTER_INVITE message to each neighbor. During the second phase, only the winning clusterhead nodes send out a CLUSTERHEAD message, followed by all accepting nodes sending a CLUSTER_ACK message. In round three all successful clusterheads reply with another CLUSTERHEAD message. These messages can be enumerated as follows:

CLUSTER_INVITE: The nodes will send this message out to all neighbors in the cluster candidate list. The length of the cluster candidate list is bounded by the MAX_CLUSTER_SIZE. Thus, in high density cases, each node will only send out MAX_CLUSTER_SIZE-1 messages. In low density cases, each node will only transmit this message to its neighbors, which has been determined to be n_{cov}. Thus, depending on network density, assuming n nodes in the network, the number of messages will be:

$$\begin{aligned} CI_{low} &= n \cdot n_{\mathrm{cov}} \\ CI_{high} &= n \cdot (MAX\_CLUSTER\_SIZE - 1) \end{aligned} \tag{6.8}$$

CLUSTERHEAD (1): The first CLUSTERHEAD message is sent by each potential winner if it has decided it is the most connected node in its immediate area. It is more convenient to investigate the number of CLUSTERHEAD messages at this point by examining the number of messages that will be received by the winning nodes.

In either case, all the winners will send this message to all the nodes on their respective candidate lists. In the low density case, we can assume that all groups form clusters of size $(n_{\text{cov}} + 1)$, which includes the winning node, as well as all of its surrounding nodes. Thus, the number of winners will be on the order of:

$$\frac{n}{n_{\text{cov}} + 1} \tag{6.9}$$

which is derived by simply dividing the total number of nodes by the cluster size. From the message reception point of view, it is unlikely that any winning node will be invited by any other winning node due to low network density. So, all nodes, except all of the winning nodes, will receive this message. This results in:

$$CH_{1,low} = n \cdot \left(1 - \frac{1}{n_{\text{cov}} + 1}\right) \tag{6.10}$$

In the high density case, though, it is possible that every node receives a message, as there will likely be winning nodes within a single hop of each other. Thus, if we assume that each node receives a message,

$$CH_{1,high} = n \tag{6.11}$$

CLUSTER_ACK/CLUSTERHEAD (2): In most cases, all nodes which reply with a CLUSTER_ACK message will receive a CLUSTERHEAD (2) message confirming that the cluster has formed. By simulation, it has been determined that, given the number of nodes in the network, approximately half will participate in final cluster formation (not including the clusterheads). Thus, for each message, in either density, the number of messages has been determined to be:

$$CA_{low} = CA_{high} = CH_{2,low} = CH_{2,high} = \frac{n}{2} \tag{6.12}$$

By combining the above derivations, an upper bound can be determined for the energy consumption due to signaling associated with the D-PEC algorithm. Given a number of nodes (n) and the degree of connectivity (n_{cov}), the number of messages M is determined to be:

$$\begin{aligned} M_{low} &= n \cdot \left\lceil \frac{n_{\text{cov}}^2 + 3n_{\text{cov}} + 1}{n_{\text{cov}} + 1} \right\rceil \\ M_{high} &= n \cdot (MAX\_CLUSTER\_SIZE + 1) \end{aligned} \tag{6.13}$$

6.3.4 Comparison to LEACH and SWE

In the LEACH algorithm, nodes are elected, in a distributed fashion, to become clusterheads on a random basis, with provisions to allow all nodes to become clusterheads during some part of the network's clustering cycle. Thus, assuming a fraction (p) of the nodes are to be clusterheads, then there will be (p^{-1}) reconfigurations to assure all nodes are allowed to become clusterheads. The algorithm for cluster formation follows a three round cycle. In the first round, all nodes that decide on being clusterheads will transmit intent across the entire network using a non-persistent CSMA protocol during which all other nodes must leave their radios on. In the second round, each non-clusterhead node selects the optimal clusterhead and replies with a request to join, again using a non-persistent CSMA protocol, during which the clusterheads must continually listen. Finally, in the third round, each clusterhead must calculate the TDMA schedule with which to communicate with the nodes, and transmit this information to the member nodes.

To determine the effect of long distance transmissions on the energy reserves, the value of γ, shown above to be the equivalent number of short distance transmissions, is used. For an example where a network area was selected to be 50 meters by 50 meters, the longest possible transmission is on the order of 70 meters. Assuming the radio specifications given above:

$$\begin{aligned} \gamma_{10m} &= 1 \\ \gamma_{50m} &\simeq 2.62 \\ \gamma_{70m} &\simeq 7.21 \end{aligned} \tag{6.14}$$

For the energy costs due to receiver power, an examination of the CSMA protocol was simulated. The goal is to determine the number of slots that must be used to receive all of the pending transmissions, including collisions. We assume k sources attempt to transmit, with an initial back-off time h (the upper bound for a uniform random variable). In the event of a collision, the nodes involved in the collision will double their back-off time (bound), and generate a new random delay. Results were determined for situations where $k = \{3,5,10,19\}$, by determining the optimal back-off bound, and then simulating the average slot delay, and number of transmissions. Table 6.1 shows these results.

k	Slot Delay	Total Transmissions
3	**8.047**	**5.274**
5	**17.479**	**9.334**
10	**46.169**	**20.401**
19	**110**	**36**

Table 6.1 Slot delay and number of transmission for k-backoff CSMA

As an sample scenario, we assume a 100 node network, with 5 cluster head nodes (randomly chosen). The field dimensions are 50 meters by 50 meters. On average, there will be 20 nodes per cluster, 1 clusterhead and 19 member nodes. The energy used in each round is given as:

Round 1: 5 Clusterheads transmit messages to the entire network with 95 nodes receiving. Using Table 6.1, there will be 9.334 transmissions and the transmission time will take 17.479 slots. Assuming a value of $\gamma = 7$, we derive:

$$\begin{aligned} E_t &= (9.334)\cdot(7)\cdot\widehat{E}_t \\ &= 67.4\cdot\widehat{E}_t \\ E_r &= (95 nodes)\cdot\left(\frac{17.5716 slots}{node}\right)\cdot\widehat{E}_r \\ &= 1,669\cdot\widehat{E}_r \end{aligned} \tag{6.15}$$

Round 2: Each non-clusterhead node will select the optimal clusterhead (based on signal quality) and reply with a request to join the cluster group. We assume, for our calculations, that one cluster will not interfere with an adjoining cluster (no CSMA interference). In this case, we have 19 nodes transmitting via the CSMA protocol (medium distance transmissions).

$$
\begin{aligned}
E_t &= (36)\cdot(3)\cdot(5clusters)\cdot \widehat{E}_t \\
&= 540\cdot \widehat{E}_t \\
E_r &= (5nodes)\cdot\left(\frac{110slots}{node}\right)\cdot \widehat{E}_r \\
&= 550\cdot \widehat{E}_r
\end{aligned}
\tag{6.16}
$$

Round 3: Each clusterhead transmits the final TDMA schedule to their member nodes. Again, assuming no CSMA interference from cluster to cluster, there will be one transmission (medium distance) and 19 receptions per cluster.

$$
\begin{aligned}
E_t &= (3)\cdot 5\cdot \widehat{E}_t = 15\cdot \widehat{E}_t \\
E_r &= 95\cdot \widehat{E}_r
\end{aligned}
\tag{6.17}
$$

This yields a total of:

$$
\begin{aligned}
E_t &= 622\cdot \widehat{E}_t \\
E_r &= 2,314\cdot \widehat{E}_r
\end{aligned}
\tag{6.18}
$$

It is likely, though, that these results will be higher as optimal assumptions were assumed. In the CSMA reception phase, for example, it is not likely that any node will know how many clusterheads are transmitting, and thus must wait a longer time to be sure that all clusterheads have reported. Also, we have assumed a mean slot delay for reception energy usage. In the 5 clusterhead case, though, the receiving nodes would have to wait approximately 55 slots for all 5 nodes to report with 99 percent assurance. In rounds two and three, we have indicated that no intra-cluster contention will occur, which will generally degrade the performance of the CSMA protocol. Furthermore, as the area of the network increases, the algorithm breaks down due to the immense transmission power required to inform the entire network of clusterhead election. Table 6.2 gives some more results of the LEACH protocol.

In the Single Winner Election (SWE) algorithm, all nodes wishing to participate in data sharing will elect to be part of a cluster. Each node assumes it will be the central node, or clusterhead, and sends this intent to all neighbors. Upon receiving this message, each other node will decide on a new clusterhead based on some metric, such as energy reserves or received target signal strength, and rebroadcast this message to all nodes. Nodes not wishing to be part of this cluster will drop these intent messages, as will nodes not wishing to update their corresponding clusterheads. Eventually, the signaling portion of this algorithm will conclude, and data sharing can begin.

	Messages Transmitted	Frame Delay
LEACH(100 nodes,50m by 50m)		
3 Clusterheads	$622 \cdot \widehat{E}_t + 1{,}538 \cdot \widehat{E}_r$	**N/A**
5 Clusterheads	$622 \cdot \widehat{E}_t + 2{,}314 \cdot \widehat{E}_r$	**N/A**
10 Clusterheads	$684 \cdot \widehat{E}_t + 4{,}630 \cdot \widehat{E}_r$	**N/A**
SWE (per cluster)		
5 Node	$10 \cdot \widehat{E}_t + 10 \cdot \widehat{E}_r$ *	**2.5**
7 Node	$20 \cdot \widehat{E}_t + 20 \cdot \widehat{E}_r$	**3.0**
D-PEC(100 nodes, 50m by 50m)		
Maximum Cluster = 5 Nodes	$600 \cdot \widehat{E}_t + 600 \cdot \widehat{E}_r$	**N/A**
Maximum Cluster = 7 Nodes	$800 \cdot \widehat{E}_t + 800 \cdot \widehat{E}_r$	**N/A**
Maximum Cluster = 9 Nodes	$1{,}000 \cdot \widehat{E}_t + 1{,}000 \cdot \widehat{E}_r$	**N/A**

(*This is per cluster, per event.)

Table 6.2 Overhead comparison of LEACH, SWE and D-PEC

Simulations of this algorithm were performed in MATLAB by generating n nodes in an area defined by the required density or node connectivity. Without loss of generality, one node was elected as a clusterhead, initiating a message surpressing all other nodes from declaring a similar intent. The cluster is allowed to form via the SWE algorithm. If enough nodes participate in the cluster, defined by the minimum cluster size, then the number of messages and the frame length delay were recorded. Results have been tabulated for maximum allowable cluster sizes of 5 and 7 nodes, with a minimum required size of 3 nodes. In reality, though, there is no method of predetermining cluster size. Table 6.2 presents these results.

The primary drawback of the SWE event-driven clustering scheme is the delay involved in cluster formation. This may present a problem for time-sensitive tracking scenarios, such a mobile tracking. Also, for frequently occurring clusters, the signaling will need to be recurrent, again with the associated delay. This, coupled with the fact that for any event multiple clusters will need to form, suggests that the energy usage will sharply rise after only a modest number of events. For example, if 3 clusters form per event, and 10 events occur, we see a factor of 30 increase over the energy used to form just 1 cluster.

6.3.5 Simulation Results

To illustrate the signaling performance of the D-PEC algorithm, simulations were performed in MATLAB. Nodes were randomly distributed in a sensor field and allowed to cluster via the D-PEC algorithm. Varied parameters included minimum and maximum cluster sizes, field area, and node density. In each scenario, the density of the field was increased, effectively increasing the average connectivity of each node. Relationships were then determined between the average signaling per node versus the average connectivity of each node. The appropriate bounds derived above are also depicted.

Figure 6.3 demonstrates the effect of the maximum cluster size on the signaling overhead. The minimum cluster size was held at 3, while the maximum cluster size varied from 5 to 9. The dashed lines represent the bounds analytically computed for the signaling overhead in the low density and high density cases. Figure 6.4 shows the independence of the results on the minimum cluster size. With the maximum cluster size held at 7, the minimum cluster size was allowed to vary from 3 to 5, showing little change in the results. Figure 6.5 shows the effect of higher density fields on the D-PEC algorithm. Note that as the density increases, the signaling overhead remains virtually constant. Figure 6.6 shows the distributed properties of the D-PEC algorithm. As the area is increased, without decreasing the density, the algorithm shows little change in the signaling per node.

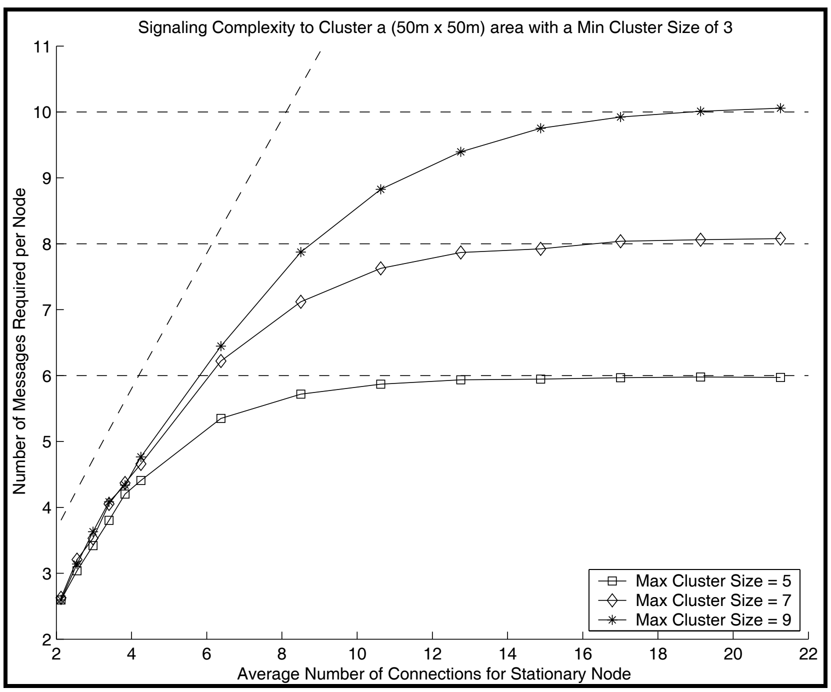


Figure 6. 3 Signaling overhead for D-PEC, varying maximum cluster size

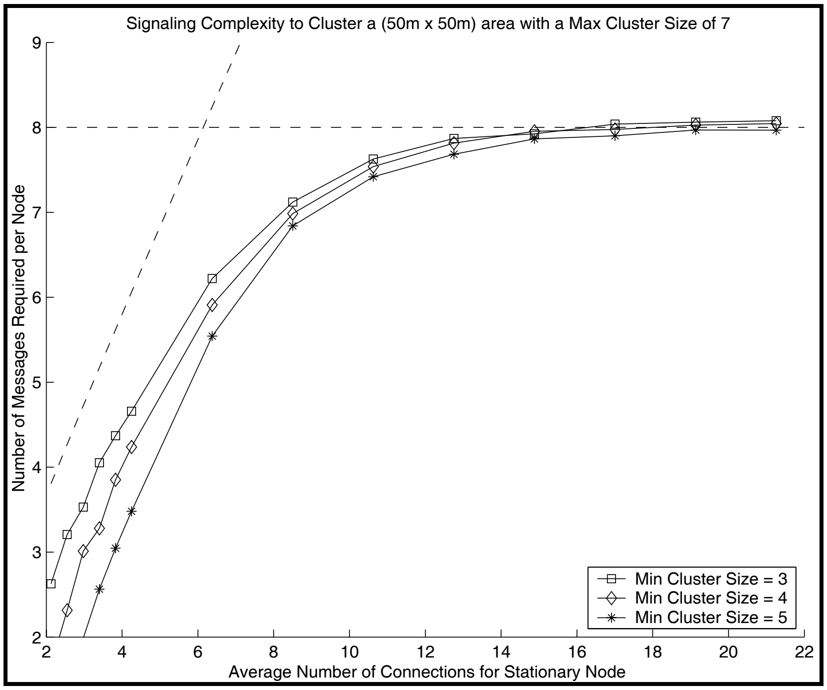


Figure 6. 4 Signaling overhead for D-PEC, varying minimum cluster size

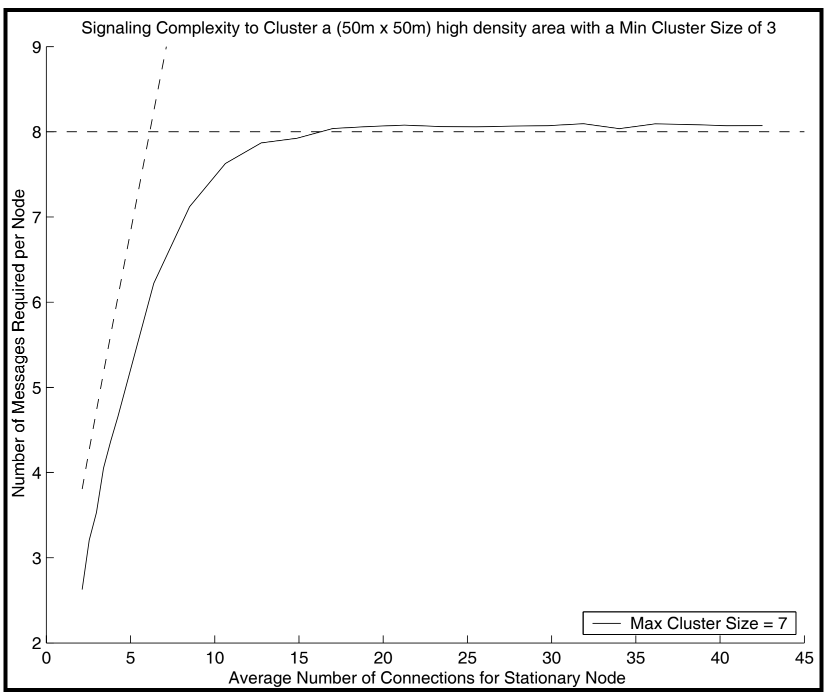


Figure 6. 5 Signaling overhead for D-PEC, varying density

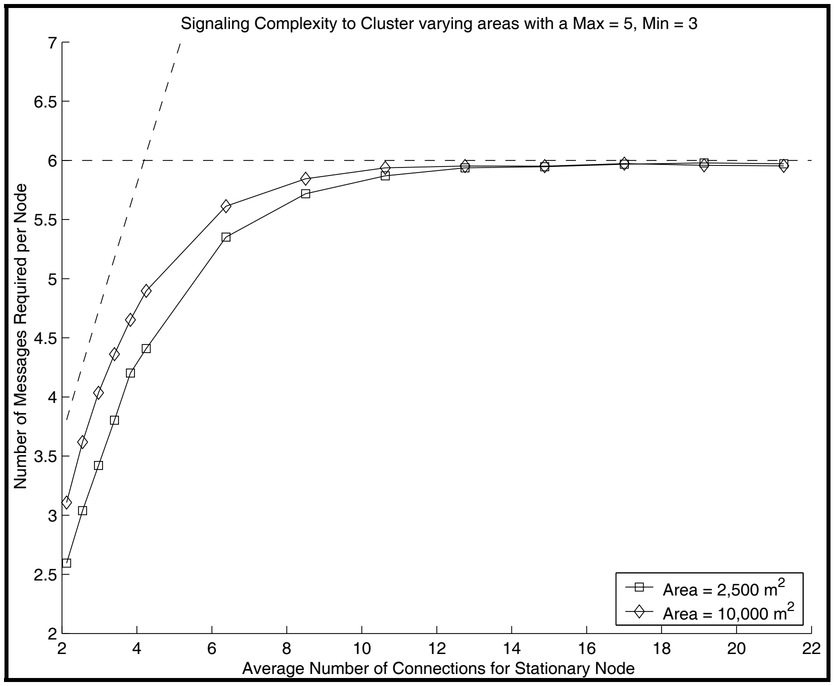


Figure 6. 6 Signaling overhead for D-PEC, varying network area

6.4 Conclusion

A new scheme to cluster a wireless sensor network in a distributed sense is presented as the D-PEC algorithm. We investigate the possibility of data sharing via clusters to effectively identify and track mobile or frequently occurring targets. By noticing that consecutive mobile events tend to be spatially correlated, a sensor network can be clustered prior to event occurrence to provide energy savings due to the signaling overhead incurred during formation.

A brief description of centralized and distributed clustering methods is provided, including hierarchical clustering (LEACH) and event driven clustering (SWE). Both are shown to cause a significant drain on energy reserves. Alternatively, D-PEC allows the sensor network to form clusters prior to event occurrence. This is a three-phase scheme which selects clusterheads based on maximum node connectivity and lowest ID. The cluster members will form one-hop (star formation) membership with the clusterhead. Furthermore, the clusters will be distributed evenly throughout the network area.

The upper bound on the signaling overhead per node is dependent only on the maximum allowable cluster size in the dense network case, and the node connectivity otherwise. D-PEC was shown to outperform competitive schemes in terms of energy usage in forming clusters. Through simulations, the effects of varying user definable parameters and network topology were given. As the density of the network increased, the signaling complexity was dependent only on the maximum allowable cluster size. Also, as the density was held constant and the network area was increased, the signaling complexity per node remained unchanged.

Chapter 7

Conclusion

We present here the novel concept of Hybrid Ad-Hoc Wireless Sensor Networks (HANETs). These networks are envisioned to consist of a large set of densely distributed stationary nodes, with the inclusion of a small subset of mobile nodes. A brief investigation of various other mobility-support wireless network architectures is presented, including Cellular Networks and MANETs. The HANET architecture suggests a new suite of protocols to account for constrained energy supplies at the stationary nodes and low degree of mobility.

We found that it is possible to reduce the energy usage at the stationary nodes by allowing for a mobile-centric link level connection protocol. By exploiting the high density of the stationary network, and thus the proximity between a mobile node to a stationary node, we were able to increase the required threshold of signal power to initiate connections. The implications of the overall rise in received signal power allowed the mobile nodes to bypass standard handoffs and avoid the use of acknowledgement messages. This reduced signaling became the prime source of energy savings. Furthermore, this was accomplished while maintaining a high quality of service, such as low outage and low bit error rate.

By exploiting the stationary wireless backbone as a routing mechanism, it was possible to avoid heavy signaling to set up routes from the mobile node to the sink node.

As the backbone is stationary, the Table Driven architecture associated with routing need only be computed once, with local routing adjustments accounting for mobility at the end nodes. We found that the stationary network was able to set up a reverse route when a connection was formed, with only minor adjustments to this route during MAC level handoffs. The combined use of the path updating and routing schemes allowed for a low packet dropping rate, a low packet delay, and a low signaling overhead as compared to various packet forwarding schemes.

Further work in the areas of MAC level and Network level protocol design involve the more sophisticated use of mobile nodes in the HANET operations. First, it may be possible to allow direct mobile to mobile connections at the MAC level, thereby avoiding any energy drain at the stationary nodes when two mobile nodes in close proximity. The mobile nodes can also act as part of the backbone network, aiding the multihop routing process. This becomes attractive when a localized set of nodes has depleted energy supplies, and a mobile node can simply move to this location and act as an information bridge to the rest of the network.

To further reduce the energy consumption at both the mobile and stationary nodes, we allow the MAC level protocol to control the node's radio operations. When connected, without the use of radio control, both nodes will need to continuously update their received signal quality with control messages. Also, while the stationary nodes are constantly transmitting pilot signals, the mobile node will need to continuously monitor the channel for future connections. This radio functionality can be allowed to occur less frequently, or even be bypassed completely, during periods of strong connections, suggesting that a handoff will not be necessary for some time. The mobile node can make a prediction as to the amount of time a connection will be valid, in a stochastic sense, to determine when to update connections or search for new connections. We have shown that by using this probabilistic scheme, the signaling energy can be reduced, while maintaining low outage probability and high connection maintenance probability.

In further work, it will be interesting to see how the radio control functions with the EAR protocol, as the EAR protocol will require continuous connection updates from the surrounding nodes to function.

In some cases, the intention of a mobile within the stationary network will not be known. Or, it may be possible that the stationary network is interested in tracking its own constituent mobile nodes. As mobility tends to be a correlated event in time and space, we have proposed a cluster scheme which allows for localized data sharing without the need for repeated cluster formation. The pre-event driven scheme has been shown to operate in a fully distributed sense, for varying densities of nodes. Furthermore, we can bound the signaling overhead associated with each node, even in high density environments, by limiting the sizes of the clusters. The clusters themselves are composed of a clusterhead node, which is the most connected node in the group, with each member node being one hop away to avoid inter-cluster multihop routing. Also, the cluster sizes are fully user definable. By eliminating the need for repeated event-driven cluster formation, we have shown a significant decrease in the signaling overhead related to clustering for the purposes of data sharing.

In further research, it may be interesting to note that prior to deployment, certain nodes can be designated as clusterheads. These nodes may be given more sophisticated radios (increasing transmission distances) and higher energy reserves (to allow for localized processing). In this scheme, the initial phases of clusterhead determination may be avoided, thereby offering a reduced energy usage due to invitation messages.

As an overall research direction, it would be interesting to note that we have assumed that the level of mobility is quite low, both in velocity and number. That is, we assume that the mobile units are slow moving robotic nodes or personnel, and are grossly outnumbered by stationary nodes. It would be interesting to investigate the breakpoints at which other protocols, such as those designed for other mobile wireless networks, would need to be used.

Appendix A

A Spatially Correlated Radio Channel Model

A.1 Introduction

As a transmitted waveform moves through free space, it becomes important to investigate the properties of the received signal, including the received signal power, at any given point. In general, the strength of the signal may be degraded when received due to many factors, most prominently being the environment. The importance of determining environmental effects may be significant if, for example, cellular transmitters need to be efficiently located or network connectivity needs to be determined. Of all of the methods which can be used to identify these environmental factors, hands-on experimental testing may be the most reliable. Unfortunately, it is usually the most expensive. Thus, statistical modeling is an attractive alternative solution.

A model is presented here which may simplify the approximations of the three major components of signal attenuation: free-space loss, shadowing (slow fading), and multipath interference (fast fading). This simplification is used to provide ease of calculation and efficiency, but will also preserve the basic statistical essence of each gain component. The derivations and implementations of the model will be demonstrated.

A.2 Model Development

As mentioned above, the net gain of the signal can be broken up into three different components: (a) free-space gain, (b) shadowing gain, and (c) multipath interference gain. Because the main focus of this model is to develop a way to simulate the losses in a communications channel by statistical analysis, it is necessary to analyze each gain component separately.

A.2.1 Free Space Gain

The first model involving radial free space loss uses an inverted exponential rule in which the absolute coordinates of the transmitter and receiver are used. For any signal path vector, the distance of the path traversed by the wave can be determined. The separation between the two objects is given as:

$$d = \sqrt{(r_x - t_x)^2 + (r_y - t_y)^2} \tag{A.1}$$

where (r_x, r_y) and (t_x, t_y) are the coordinates of the receiver and the transmitter, respectively. The gain, then, of the system due to free space loss is given by

$$G_{\text{distance}} = \frac{1}{d^n} \tag{A.2}$$

with the gain equaling one when the distance drops below one meter.

A.2.2 Shadowing Gain

Unlike the free space gain, the shadowing gain is not easily computed deterministically. To obtain exact values, it would be necessary to develop a correlation function including every point on the environmental layout. As explained above, though, the shadowing gain can be thought of as being a combination of a series of many shadowing gains over smaller distances. In general, any value of p, the shadowing constant, greater than zero represents a net gain, while any value less than zero represents an attenuation. Also, the gain due to shadowing is not dependent on the distance between the transmitter/receiver pair, but on their relative spatial location to surrounding objects.

In a statistical sense, the environment layout can be considered to have objects in random positions with relation to the transmitter and receiver. Also, the shadowing constants for each object can be considered to be random, allowing a representation by a Gaussian distribution. By experimentation, it can be shown that different geometries imply different variances used to characterize the shadowing constant.

We propose that each object (transmitter or receiver) will have a shadowing constant associated with it, depending on location. Assuming some degree of object mobility, it may be required to provide some level of spatial correlation for the shadowing measurements. Models such as [17] have been suggested, providing an autoregressive spatial correlating method. This, though, does not account for the situations where a mobile object returns to a similar location, or the velocity of the mobile is varied. To provide for every contingency would force each possible point in the "world" to have a shadowing value associated with it, with each being spatially correlated in two dimensions. We propose that each object is located within a two-dimensional grid space, with each grid point being assigned a shadowing constant. The spacing of the grid points is related to the magnitude of the spatial correlation. To determine the shadowing constant at any given point, the four surrounding points can be used via an interpolation method. In this way, along any trajectory, shadowing values can be determined and combined to result in the net shadowing constant value, or gain, for any given configuration.

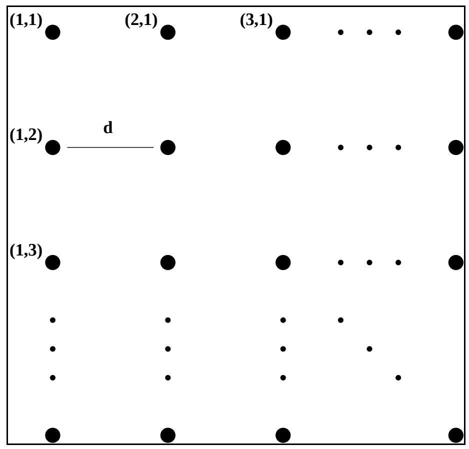


Figure A. 1 Shadow grid example

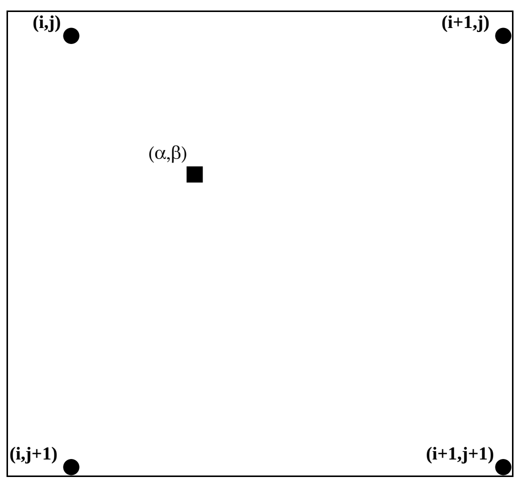


Figure A. 2 Position of shadowing interpolation point

To accomplish this model, the first step is to create a discrete, theoretical space onto which the random shadowing constants would be arranged. Spatial models require the grid to be bounded by approximately 1000 by 1000 points. In these scenarios, each grid point was predetermined by a Gaussian random variable with a standard deviation of 8 dB. In this manner, the layout would remain constant, allowing an object which returns to a previous position to find the same shadowing gain. Figure A.1 depicts this virtual grid, with the spacing, d , representing the distance between new shadowing grid points. Thus, translating an object from its real space to the theoretical grid space is accomplished by dividing its absolute coordinate pair by the grid spacing, giving its relative grid location. Now, each object is placed on the grid, surrounded by four, equally spaced, Gaussian distributed shadowing constants (assuming the spacing in both directions is equal, which may not be the case), as in figure A.2.

It is important to realize the effect of the grid spacing on the performance of the model. Any point in this theoretical grid has associated with it a random shadowing value. Recalculating new values at each motion instance would provide completely uncorrelated shadowing values. As the receiver, or transmitter, in the system moves slightly in any direction, the new shadowing values should be related to previously computed values. Thus, the four surrounding points are used to form an interpolated value. As the object moves around the grid, therefore, these values remain continuous. When the object crosses a boundary, as shown in figure A.3, two new points arise to accommodate the new quadrant. The grid spacing, therefore, determines the correlation of the shadowing values. If, for instance, the object reaches a new quadrant at each sampling instance, each value would have a low correlation to previous values.

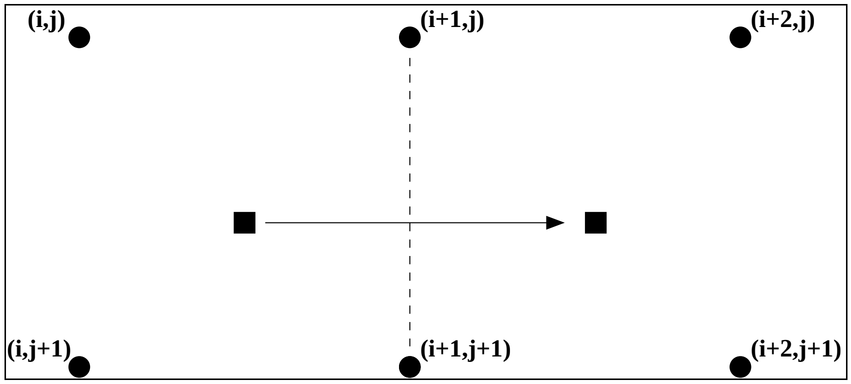


Figure A. 3 Shadowing grid boundary crossing

To simplify the situation, as the results will be statistical, it is only necessary to calculate partial shadowing components for the transmitter/receiver pair. It is important, therefore, that during the interpolation of the shadowing point, the statistics of the Gaussian random variables is preserved. At any point in the plane, there are four shadowing points which form a box around the object, as shown in figure A.4. As each of the grid points are normally distributed with a specified standard deviation, σ_g, the resulting interpolated point must also yield the same standard deviation. In figure A.4, the four surrounding points are given as g_i, where $i = \{1,2,3,4\}$. A perpendicular line extending upwards from the object, point p, is located between points g_1 and g_2, and is called p_{up}. Of key importance are the weights used in combining the two normal variables. They must be chosen to maintain the variance of the interpolated point. As the distance between p_{up} and g_1 is given in figure A.4 as x, and the grid spacing is d, the following derivation can take place. The value of p_{up} is as follows:

$$p_{up} = (w_1 \cdot g_1) + (w_2 \cdot g_2) \tag{A.3}$$

$$\sigma_p^2 = \left(w_1^2 \cdot \sigma_g^2\right) + \left(w_2^2 \cdot \sigma_g^2\right) \tag{A.4}$$

where σ_p^2 and σ_g^2 are the variances of the new point and the Gaussian grid points, respectively [29]. As these variances must be equal, we obtain:

$$w_1^2 + w_2^2 = 1 \tag{A.5}$$

Next, we notice that as the distance x approaches zero, w_1 should approach one, allowing the entire weight to be over the first Gaussian grid point, with a similar result for x approaching d on w_2. Thus, the following forms for the weights can be assumed:

$$\begin{aligned} w_1 &= \left(\frac{d-x}{d}\right) f(x) \\ w_2 &= \left(\frac{x}{d}\right) f(x) \end{aligned} \tag{A.6}$$

Combining the equations in A.5 and A.6, we find the result for $f(x)$, and thus for p_{up}.

$$p_{up} = \frac{\left(\frac{d-x}{d}\right)\cdot g_1 + \left(\frac{x}{d}\right)\cdot g_2}{\sqrt{\left(\frac{d-x}{d}\right)^2 + \left(\frac{x}{d}\right)^2}} = \frac{(d-x)\cdot g_1 + (x)\cdot g_2}{\sqrt{(d-x)^2 + (x)^2}} \tag{A.7}$$

By the above derivation, the variance of p_{up} is equal to the variance of both g_1 and g_2. Now, given that a similar derivation can be performed for p_{down}, and that y and h represent the distances used to combine p_{up} and p_{down}, an extended expression for p utilizing all four points is given as:

$$p = \frac{(d-x)\cdot(h-y)\cdot g_1 + (x)\cdot(h-y)\cdot g_2 + (d-x)\cdot(y)\cdot g_3 + (x)\cdot(y)\cdot g_4}{\sqrt{\left((d-x)^2 + (x)^2\right)\cdot\left((h-y)^2 + (y)^2\right)}} \tag{A.8}$$

Again, this value preserves the variance of each Gaussian random variable.

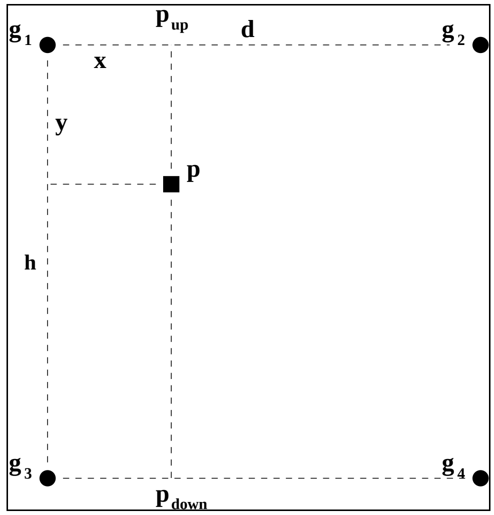


Figure A. 4 Shadowing grid interpolation

Now, calculating the shadowing value for the transmitter/receiver pair, namely p_t and p_r, is required to determine the net shadowing gain. Again, to preserve the statistics of the model, the expression is:

$$p = \frac{1}{\sqrt{2}}\left(p_t + p_r\right) \tag{A.9}$$

This value of p is the shadowing gain given in units of dB. Finally, the absolute shadowing gain can be found by

$$G_{\text{shadow}} = 10^{(0.1)p} . \tag{A.10}$$

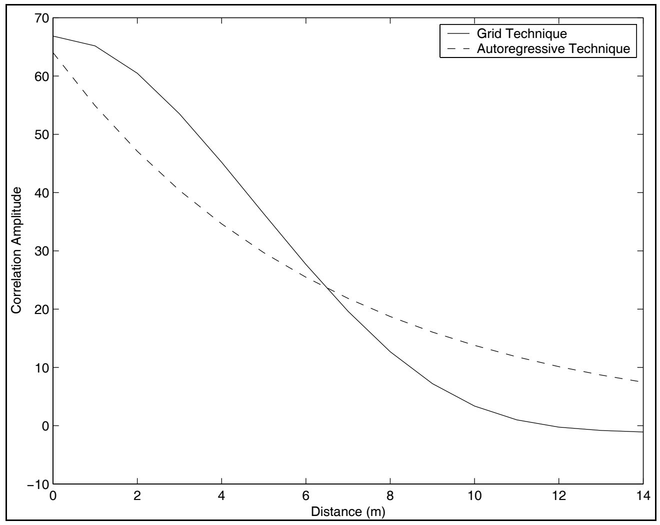


Figure A. 5 Comparison to correlated shadowing (5 meters)

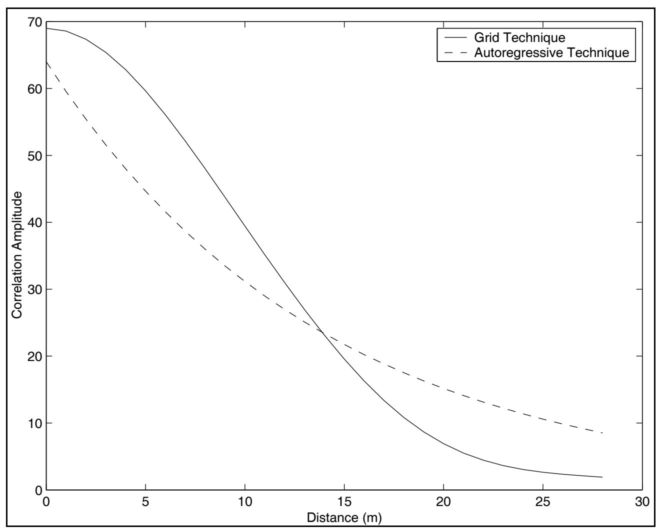


Figure A. 6 Comparison to correlated shadowing (10 meters)

The correlation properties of the shadowing grid technique are compared with that presented in [17]. Simulations were performed in MATLAB, with mobile objects traveling through virtual grids with grid spacings varying from 5 meters to 10 meters. The spatial correlation from the measured shadowing values was determined. Figure A.5 shows the results for a 5 meter grid separation. The trend for the grid technique is similar to that of the auto-regressive technique. In fact, numerically, as the mobile object travels at a zero degree angle to the grid spacing, a correlation distance of about 0.9 times the grid spacing is found. For 45 degree motion, the correlation approaches approximately 1.3 times the grid spacing. As the mobile was allowed to freely move about the grid, the net correlation distance was similar to the grad spacing. Figure A.6 gives the same results for a 10 meter separation of grid points.

This model yields continuous shadowing values for small movements of transmitters and receivers, providing sufficiently accurate results for stochastic processing.

A.2.3 Multipath Interference Gain

The final model developed is that of the multipath interference gain. Unlike the shadowing procedure, the multipath interference cannot assume a widely spaced grid-like configuration, due to the short length spatial correlation (on the order of fractions of wavelengths, centimeters). The shadowing gain, on the other hand, will change relatively slowly with movement on the order of meters. Thus, a new method must be developed.

The multipath model can be envisioned as a combination of many partial waves at the receiver. In particular, for any k-th partial wave, its representation is given as $a_k e^{j\theta_k}$. This k-th wave is characterized by its time delay in reaching the receiver. For the special case where the delay times τ_k are discrete, any partial wave arriving within a time interval described by

$$\begin{aligned}&[\tau_k, \tau_{k+1}) \\ &\tau_k < \tau_{k+1}\end{aligned} \tag{A.11}$$

will be labeled as part of the k-th ray, and combined with all other k-th rays. Thus, the combination of the k-th amplitude and phase is in itself a combination of many partial waves. The final recombined wave is given as

$$h(\tau) = \sum_i a_i e^{j\theta_i} \delta(t - \tau_0 - \Delta\tau_i)\ . \tag{A.12}$$

At this point, for large values of k, the central limit theorem can be used to approximate the statistics of a_k and θ_k. In particular, looking at the in-phase and quadrature components, each takes on the characteristics of a gaussian random variable. Thus, the implementation would be to generate two new independent unit variance Gaussian random variables, one for the in-phase component and one for the quadrature component, whenever the object moves on the order of a wavelength.

Generating new Gaussian values continuously, though, provides no correlation between previously obtained values of multipath gain. To overcome this, an autocorrelation function is used. As this is a discrete windowing filter, the following form can be used:

$$h(t) = \sum_{i=1}^{n} w_i \delta(t - t_i - \tau) \tag{A.13}$$

where $h(t)$ is an n-point discrete filter and t_i is the spacing between random variable generation. Again, by probability theory, the weights w_i should be chosen to allow the newly calculated points to have unit variance. The final value takes on the form of a convolutional difference equation:

$$m_i = w_1 \cdot g_i + w_2 \cdot g_{i\text{-}1} + w_3 \cdot g_{i-2} + \cdots + w_{n-1} \cdot g_{i-n+2} + w_n \cdot g_{i-n+1} \tag{A.14}$$

where m_i is the i-th calculated multipath interference gain, in any component (in-phase or quadrature). To preserve the value of the unit variance, the following must hold:

$$\sum_{i=1}^{n} w_i^2 = 1 \tag{A.15}$$

This property allows the filter to have a power value of one [29]. At this point, any filter length can be used, with common choices being a raised cosine filter or a Gaussian shaped filter. For our implementation, a 32-point Gaussian filter is developed. The spacing between the delta functions in the filter are user-definable as the frequency of calculation. At this frequency, a new multipath interference gain, as well as a free-space gain and a shadowing gain, in computed. The convolution takes on the following form:

$$y(t) = \sum_{i=1}^{32} x(t-i) \cdot w_i \tag{A.16}$$

Here, the values of the weights are defined as:

$$w_i = A \cdot \frac{1}{2} e^{-\left(\frac{i-15.5}{4}\right)^2} \tag{A.17}$$

where the value of A is the normalization factor which forces the unit power constraint. During implementation, an initial bank of 32 Gaussian points will need to be computed for use in the 32-tap difference equation.

Appendix B

Energy Costs Associated with Long Distance Transmissions

When comparing two different protocols used in wireless networks, it is common to discuss differences in signaling overhead and energy utilization. This task is as simple as counting messages in short distance multihop connection oriented networks. Some configurations, though, require that a transmitter shares information with receivers at relatively large distances, thus incurring a higher power usage due to amplification costs. The goal of this section, therefore, is to draw a numerical comparison between long distance and short distance transmissions.

B.1 Radio Specifications

We define the transmission power usage and the reception power usage as follows:

$$\begin{aligned} P_t &= P_{diss} + P_{amp} \\ P_r &= P_{diss} \end{aligned} \tag{B.1}$$

where P_{diss} represents the power dissipated by the radio when powered on. The value of P_{amp} is simply the extra power required to achieve a given SNR at the reception point. Similarly, we can define the energy per bit during energy dissipation and amplification as

$E_{diss}(J/bit)$ and $E_{amp}(J/bit/m^{\alpha})$, assuming an $\alpha-$order exponential free-space power falloff. Then, if the transmission time (slot length) is assumed to be $T(\sec)$, and the bandwidth of the signal is $B(bits/\sec)$, the total energy used for transmissions and receptions is:

$$\begin{aligned} E_t &= T \cdot B \cdot E_{diss} + T \cdot B \cdot r^{\alpha} \cdot E_{amp} \\ E_r &= T \cdot B \cdot E_{diss} \end{aligned} \tag{B.2}$$

In Chapter 5, we derived an expression for the received signal SNR as:

$$SNR_{rec} = \left(\frac{B'}{B}\right) \cdot \left(\frac{P_t}{P_t'}\right) \cdot \left(\frac{r'}{r}\right)^{\alpha} \cdot \left(\frac{\Phi}{\overline{\Phi}}\right) \cdot SNR_{req} \tag{B.3}$$

where B is the signal bandwidth, P_t is the transmission power, and r is the transmission distance. To incorporate environmental effects such as shadowing and multipath, a random variable a used, which is user describable, as Φ, with mean $\overline{\Phi}$. The values of B', P_t', and r' represent the bandwidth, power, and distance values required to force the received signal SNR to be equal to the required signal SNR for reception when the random variable equals its mean. For the purposes of the discussion here, we will assume that the bandwidth for near field transmissions and far field transmissions is the same. Also, we assume no random environmental factors will effect the signal. Thus, equation B.3 becomes:

$$SNR_{rec} = \left(\frac{P_t}{P_t'}\right) \cdot \left(\frac{r'}{r}\right)^{\alpha} \cdot SNR_{req} \tag{B.4}$$

If we assume that power control methods are used which guarantee that the received signal SNR in the near and far field cases will both be precisely equal to the required SNR, then we see that:

$$P_t = \left(\frac{r}{r^{'}} \right)^{\alpha} \cdot P_t^{'} \tag{B.5}$$

where the values of $r^{'}$ and $P_t^{'}$ are chosen to represent the values required in the near field case. If the dissipation power is the same for the near and far field cases, then the total transmission power for both cases is found by equation B.1. In general environments, sensors will employ a fade margin to overcome slow fading, or shadowing, to guarantee a low outage probability. This boost of power arises as a multiplicative factor to the amplification power as (reference power):

$$P_t^{'} = P_{diss} + P_{amp}^{'} \cdot fm \tag{B.6}$$

Substituting this information into equation B.1, we get:

$$E = T \cdot P_{diss} + T \cdot \left(\frac{P_{amp}^{'} \cdot fm}{r^{'\alpha}} \right) \cdot r^{\alpha} \tag{B.7}$$

To find the ratio of long distance energy consumption to short distance consumption, we define:

$$\gamma = \frac{E}{E^{'}} = \frac{T \cdot P_{diss} + T \cdot \left(\frac{P_{amp}^{'} \cdot fm}{r^{'\alpha}} \right) \cdot r^{\alpha}}{T \cdot P_{diss} + T \cdot P_{amp}^{'} \cdot fm} \tag{B.8}$$

Simplifying, we find that:

$$\gamma = \frac{\Delta + \rho^{\alpha}}{\Delta + 1} \tag{B.9}$$

where

$$\Delta = \frac{P_{diss}}{P'_{amp} \cdot fm}$$
$$\rho = \frac{r}{r'} \qquad \text{(B.10)}$$

The factor γ, therefore, is the number of short distance transmissions of distance r' per long distance transmission of distance r.

As an example, we can assume the following radio specifications:

$$E_{diss} = 50nJ / bit$$
$$E_{amp} = 0.0013 pJ / bit / m^4 \qquad \text{(B.11)}$$

Assuming a short distance transmission of 10 meters, and fade margin of 10 dB,

$$\Delta = \frac{P_{diss}}{P'_{amp} \cdot fm} = \frac{50nJ / bit}{130 pJ / bit} = 385 \qquad \text{(B.12)}$$

For a 100 meter transmission, we find that:

$$\gamma = \frac{\Delta + \rho^{\alpha}}{\Delta + 1} = \frac{385 + \left(\frac{100}{10}\right)^4}{385 + 1} = 26.9 \qquad \text{(B.13)}$$

Thus, in the energy required to transmit a single message 100 meters, approximately 27 short distance transmissions of 10 meters can be made.

Bibliography

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, "Wireless Sensor Networks: A Survey," *Computer Networks*, vol. 38, no. 4, Elsevier, 15 March 2002, p. 393-422.

[2] P. Basu, N. Khan, T. Little, "A Mobility Based Metric for Clustering in Mobile Ad Hoc Networks," *Proc. 21st Int. Conf. on Dist. Comp. Sys. Workshops*, Mesa, AZ, USA, April 2001, p. 413-18.

[3] F. Bennet *et al.*, "Piconet: Embedded Mobile Networking," *IEEE Personal Communications*, vol. 4, no. 5, Oct. 1997, p. 8-15.

[4] L. Blazevic, S. Giordano and J.-Y. Le Boudec, "Self organized terminode routing," *Cluster Computing*, vol. 5, no. 2, 2002, p. 205-18.

[5] J. C. Chen, K. Yao, R. E. Hudson, "Source localization and beamforming," *IEEE Signal Processing Magazine*, vol. 19, no. 2, March 2002, p. 30-9.

[6] P. Chen, "A Non-Line-of-Sight Error Mitigation Algorithm in Location Estimation," *1999 IEEE Wireless Communications and Networking Conference*, New Orleans, LA, 1999, p.316-20.

[7] R. Chen *et al.*, "A Distributed Algorithm Enhancing the Connection Quality and Bandwidth Utilization in Wireless Cellular Networks," *Proc. 2001 Intl. Conf. on Comp. Networks and Mobile Comp.*, Los Alamitos, CA, 2001, p.79-84.

[8] M. S. Corson, J. P. Macker, G. H. Cirincione, "Internet-Based Mobile Ad Hoc Networking," *IEEE Internet Computing*, vol. 3, no. 4, July-Aug. 1999, p. 63-70.

[9] B. Crow *et al.*, "IEEE 802.11 Wireless Local Area Networks," *IEEE Communications Magazine*, vol. 35, no. 9, Sept. 1997, p. 116-26.

[10] B. Das, R. Sivakumar, and V. Bharghavan, "Routing in Ad Hoc Networks using a Spline," *Proc. IEEE ICCCN'97*, 1997.

[11] R. Dube *et al.*, "Signal Stability Based Adaptive Routing (SSA) for Ad-Hoc Mobile Networks," *IEEE Personal Communications*, Feb. 1997, p. 36-45.

[12] A. Ephremides, J. E. Wieselthier and D. J. Baker, "A Design Concept for Reliable Mobile Radio Networks with Frequency Hopping Signaling," *Proc. IEEE 75*, 1987, p. 56-73.

[13] D. Estrin, R. Govindan, J. Heidemann, S. Kumar, "Next Century Challenges: Scalable Coordination in Sensor Networks," *MobiCom'99, Proc. Fifth Ann. ACM/IEEE Int. Conf. on Mob. Comp. and Net.*, Seattle, WA, 15-20 Aug. 1999, p. 263-70.

[14] J. L. Gao, "Energy Efficient Routing for Wireless Sensor Networks," Ph.D. dissertation, Dept. of Elec. Eng., UCLA, June 2000.

[15] M. Gerla and J. Tsai, "Multicluster, Mobile, Multimedia Radio Network," *Wireless Networks*, vol. 1, no. 3, 1995, p. 255-65.

[16] D. J. Goodman, "Trends in Cellular and Cordless Communications," *IEEE Communications Magazine*, vol. 29, no. 6, June 1991. p. 31-40.

[17] M. Gudmundson, "Correlation Model for Shadow Fading in Mobile Radio Systems," *Electronic Letters*, vol. 27, no. 23, 7 Nov. 1991, p. 2145-6.

[18] W. R. Heinzelman, A. Chandrakasan, H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks**,"** *Proc. of the 33rd Annual Hawaii Int. Conf. on Sys. Sciences*, vol. 2, Maui, HI, USA, Jan. 2000.

[19] M. Hellebrandt and R. Mathar, "Location Tracking of Mobiles in Cellular Radio Networks," *IEEE Trans. on Vehicular Tech.*, vol. 48, no. 5, Sept. 1999, p. 1558-62.

[20] http://may.cs.ucla.edu/projects/parsec

[21] Internet Engineering Task Force Manet Working Group Charter, http://www.ietf.org/html.charters/mobileip-charter.html.

[22] A. Iwata *et al.***, "**Scalable Routing Strategies for Ad Hoc Wireless Networks," *IEEE Journal on Selected Areas in Comm.*, vol. 17, no. 8, Aug. 1999, p. 1369-79.

[23] D. B. Johnson and D. A. Maltz, "Dynamic Source Routing in Ad-Hoc Wireless Networks," *Mobile Computing*, 1996, p. 153-81.

[24] D. K. Kim and D. K. Sung, **"**Characterization of Soft Handoff in CDMA Systems," *IEEE Trans. on Vehicular Tech.*, vol. 48, no.4, July 1999, p. 1195-202.

[25] Y. Ko and N. Vaidya, "Location-Aided Routing in Mobile Ad-Hoc Networks," *Proc. 4th Ann. ACM/IEEE Int. Conf. on Mob. Comp. and Net. '98*, Dallas, TX, Aug. 1998.

[26] R. O. LaMaire, A. Krishna, P. Bhagwat, and J. Panian, **"**Wireless LANs and Mobile Networking: Standards and Future Directions," *IEEE Communications Magazine*, vol. 34, no. 8, Aug. 1996, p. 86-94.

[27] S. Lee, W. Su, and M. Gerla, "Wireless Ad Hoc Multicast Routing with Mobility Prediction," *Mobile Networks and Applications*, ACM, 2001, p.351-60.

[28] W. Lee, *Mobile Cellular Communications Systems*, McGraw-Hill Book Company, 1989.

[29] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, Addison-Wesley Publishing Co., 1994

[30] B. Liang and Z. Haas, “Predictive Distance-Based Mobility Management for PCS Networks,” *IEEE INFOCOM '99, Conference on Computer Communications*, New York, NY, 1999, p.1377-84.

[31] L. Litwin, “The Medium Access Control Sublayer,” *IEEE Potentials*, vol. 20, no. 4, Oct.-Nov. 2001, p. 30-4.

[32] A. O. Mahajan, A. J. Dadej and K. V.Lever, “A Comparison of Channel-Scanning Schemes for Distributed Network Formation and Reconfiguration,” *Wireless Networks*, vol. 4, no. 4, 1998, p. 341-51.

[33] O. C. Mantel, N. Scully and A. Mawira, “Radio aspects of hybrid wireless ad-hoc networks,” *IEEE VTS 53rd Vehicular Technology Conference*, Rhodes, Greece, Spring 2001, vol. 2, p. 1139-43.

[34] S. Murthy and J. J. Garcia-Luna-Aceves, “An Efficient Routing Protocol for Wireless Networks,” *ACM Mobile Networks and App. J.*, Special Issue on Routing in Mobile Communications Networks, Oct. 1996, p. 183-97.

[35] J. E. Padgett, C. G. Gunther, T. Hattori, “Overview of Wireless Personal Communications,” *IEEE Communications Magazine*, vol. 33, no. 1, Jan. 1995. p. 28-41.

[36] A. K. Parekh, “Selecting Routers in Ad-Hoc Wireless Networks,” *ITS*, 1994.

[37] V. D. Park and M. S. Corson, "A Highly Adaptive Distributed Routing Algorithm for Mobile Wireless Networks," *Proc. INFOCOM '97*, April 1997.

[38] G. Pei and M. Gerla, "Mobility management for hierarchical wireless networks," *Mobile Networks and Apps.*, vol. 6, no. 4, 2001, p. 331-7.

[39] C. E. Perkins and P. Bhagwat, "Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers," *Comp. Commun. Rev.*, Oct. 1994, p. 234-44.

[40] C. E. Perkins and E. M. Royer, "Ad-Hoc On-Demand Distance Vector Routing," *Proc. 2nd IEEE Wksp. Mobile Comp. Sys. And Apps.*, Feb. 1999, p. 90-100.

[41] M. R. Perlman and Z. J. Haas, "Determining the Optimal Configuration for the Zone Routing Protocol," *IEEE JSAC*, 1999.

[42] G. J. Pottie, "Wireless Sensor Networks," *1998 Information Theory Workshop*, Killarney, Ireland, 1998, p. 139-140.

[43] G. J. Pottie and W. J. Kaiser, "Wireless Integrated Network Sensors," *Commun. ACM*, vol. 43, no. 5, May 2000, p. 51-58.

[44] V. Raghunathan, C. Schurgers, Sung Park; M. B. Srivastava, "Energy-aware wireless microsensor networks," *IEEE Signal Processing Magazine*, vol. 19, no. 2, IEEE, March 2002, p. 40-50.

[45] E. M. Royer, C.-K. Toh, "A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks," *IEEE Personal Communications*, vol. 6, no. 2, April 1999, p. 46-55.

[46] S. Saliga, “An Introduction to IEEE 802.11 Wireless LANs,” *2000 IEEE RFIC Symp. Digest of Papers*, Boston, MA: 2000, p. 11-14.

[47] X. Shen and J. W. Mark, “Mobility information for resource management in wireless ATM networks,” *Computer Networks*, vol. 31, no. 9-10, May 1999, p. 1049-62.

[48] K. Shoarinejad, “On Stochastic Decentralized Systems in Communications and Control,” PhD Thesis, Electrical Engineering, UCLA, 2001.

[49] S. Singh and C. S. Raghavendra, “PAMAS-power aware multi-access protocol with signalling for ad hoc networks,” *Computer Comm. Review*, vol. 28, no. 3, July 1998, p. 5-26.

[50] S. Singh, M. Woo, and C. Raghavendra, “Power-Aware Routing in Mobile Ad Hoc Networks,” *Proc. of the 4th Annual ACM/IEEE Int. Conf. on Mob. Comp. and Net. (MobiCom)*, Dallas, TX, Oct. 1998.

[51] K. Sohrabi, J. Gao, V. Ailawadhi and G. Pottie, “A Self Organizing Wireless Sensor Network,” *Proc. 39th Annual Allerton Conf. Communications, Control, and Comp.*, Urbana, IL, Oct. 1999.

[52] K. Sohrabi, J. Gao, V. Ailawadhi and G. Pottie, “Protocols for Self-Organization of a Wireless Sensor Network,” *IEEE Personal Communications*, vol. 7, no. 5, Oct. 2000, p. 16-27.

[53] G. L. Stuber, *Principles of Mobile Communications*, Norwell, MA: Kluwer Academic Publishers, 1996.

[54] C.-K. Toh, "A Novel Distributed Routing Protocol to Support Ad-Hoc Mobile Computing," *Proc. 1996 IEEE 15th Annual Intl. Pheonix Conf. Comp. and Commun.*, March 1996, p. 480-86.

[55] A. Viterbi *et al*., "Soft Handoff Extends CDMA Cell Coverage and Increases Reverse Link Capacity," *IEEE Journal on Selected Areas in Comm.*, vol. 12, no. 8, Oct. 1994, p. 1281-8.

[56] J. Wen and S. Lin, "An Uplink Short-Term Fading Prediction-Based Power Control Method for DS/CDMA Cellular Radio Systems," *Wireless Personal Communications*, vol.15, no.3, Dec. 2000, p. 221-37.

[57] D. Wong and T. Lim, "Soft Handoffs in CDMA Mobile Systems," *IEEE Personal Communications*, vol. 4, no. 6, Dec. 1997, p. 6-17.

[58] L. Yip, J. C. Chen, V. Ailawadhi, R. E. Hudson, K. Yao, and G. J. Pottie, "Graphical Integrated Multi-Sensor Simulator for a Wireless Sensor Network," *Proc. ASC*, *Advanced Sensor Consortium*, March 2001, p. 31-35.

[59] R. Yuan *et al*., "A signaling and control architecture for mobility support in wireless ATM networks," *Journal of Special Topics in Mobile Networks and Applications (MONET),* vol. 1, no. 3, Dec. 1996, p. 287-98.