Gateways between ad hoc and other networks

NGUYEN MANH HA

KTH Information and Communication Technology

Master of Science Thesis
Stockholm, Sweden 2007

COS/CCS 2007-06

Gateways between ad hoc and other networks

Nguyen Manh Ha

23RD FEBRUARY 2007

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science with a major in Information Technology, specialization in Internetworking

SUPERVISOR & EXAMINER:

PROFESSOR GERALD Q. MAGUIRE JR.

Department of Communication System
School of Information and Communication Technology
KTH, Royal Institute of Technology
Stockholm, Sweden

ABSTRACT

Multi-hop wireless ad hoc wireless networks have no fixed network infrastructure. Such a network consists of multiple nodes that maintain network connectivity through wireless links. Additionally, these nodes may be mobile and thus the topology of the network may change with time. It will be useful if the nodes in this network could communicate with the Internet; this can be done via gateways which in turn inter-connect to the Internet.

This functionality requires that the nodes in the ad hoc network to discover the gateway, using a gateway discovery protocol. However, a limiting factor (particularly for mobile nodes) is suing their limited energy supply provided by batteries. In order to understand the potential effect this thesis considers two key areas: internetworking between a multi-hop mobile wireless ad hoc network and the Internet and the energy utilization as a function of number of gateways and the mobility pattern of nodes.

Using simulation on various mobility patterns and networks density scenarios, we show that increase the number of gateways in ad hoc network significantly improves the power efficiency of mobile node and therefore prevent network partition due to death nodes. The thesis also discusses about the impact of different environment and mobility patterns on the power consumption of mobile nodes which is a very important factor in the building and deployment of the cost-effective high performance wireless ad hoc networks.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

First and foremost, I would like to offer my special thanks and acknowledgment to my supervisor Professor Gerald Maguire for his excellent support and encouragement. He has been always there to answer and give me precious advice and comments on my study.

I am also grateful to Professor Ahmed Helmy at University of Southern California and Ali Hamidian at Lund University for helping me with the simulation and the thesis works. The thesis has benefited from many contributions and suggestions from ns2-users community, their contribution are really appreciated.

Most of all, I would like to thank my parents for support me and I would like to express my love to my girl friend Le Thi Thu Hang for our lovely little monkey.

GLOSSARY

WLAN Wireless Local Area Network

CSMA/CA Career Sense Multiple Access/ Collision Avoidance

BSS Basic Service Set

IBSS Independent Basic Service Set

DCF Distributed Coordination Function

PCF Point Coordination Function

SIFS Short Inter-Frame Space

DIFS DCF Inter-Frame Spacing

EIFS Extended Inter-Frame Spacing

ACK Acknowledgment

SIFS Short Inter-Frame Spacing

RPGM Reference Point Group Mobility

MANET Multi-hop Mobile Ad hoc Networks

MSDU MAC Service Data Unit

MMPDU Multi MAC Protocol Data Unit

TIM Traffic Indication Message

ATIM Announcement Traffic Indication Message

CHAPTER 1: INTRODUCTION

There have been great advances since the first invention of the wireless networks. Nowadays, many people expect to be connected at anytime, anywhere, and in anyplace. Such networks are very useful in both daily life and in emergency situations. The price of the equipment and its installation are decreasing allow wireless networks become even more popular. Although the advantages and convenience of wireless network, people always desire more than that. Most of the mobile equipments that form the wireless networks (mobile node) are rely on the limited battery power that make limit in the usage time. Longer battery life is desirable, but not always practical, affordable, or achievable.

Lowering energy consumption is a key goal in many multi-hop wireless ad hoc networking environments, especially when the individual nodes of the network are battery powered. These requirements have become increasingly important for new generations of mobile computing devices (such as Personal Digital Assistant (PDAs), laptops, and cellular phones) because the energy density achievable in batteries has grown only at a linear rate, while processing power and storage capacity have both grown exponentially. As a consequence of these technological trends, many wireless-enabled devices are now primarily energy-constrained; while they possess the ability to run many sophisticated multimedia networked applications, their operational lifetime between recharges is often short. In addition, the energy consumed in communication by the radio interfaces is often higher than, or at least comparable to, the computational energy consumed by the processor.

The effective total transmission energy, which includes the energy spent in potential retransmissions consumed per packet, is the proper metric for reliable, energy-efficient communications. The maximum and minimum of energy of candidate nodes is dependent on the number of gateway and mobility pattern of mobile nodes, since they directly affect the energy utilized in changing their immediate hop path to get to the desired external destination. Analysis of the interplay between the numbers of gateway and mobility patterns of mobile node reveals several key results. These results will be described in section 4.1 and 4.2.

The remainder of this thesis is organized as follows: chapter 2 introduces different approaches to power control in Mobile ad hoc networks (MANET). Studies about Internet connectivity for ad hoc network are briefly discussed in presented 2.2. Energy consumption model for Internet connectivity in MANET are discussed in section 2.3. Chapter 3 describes the simulation scenario and different mobility patterns. Session 4.1 discuss the effect of changing the number of gateways. The different in power energy consumption of nodes under many mobility patterns in MANET are presented and commented in section 4.2. Finally, chapter 5 summarizes and concludes the thesis.

CHAPTER 2: BACKGROUND

2.1 Different approaches to power control in ad hoc networks

Two of the most important goals in designing ad hoc networks are to provide high throughput and to lower the energy requirements of the nodes. Power saving strategies can be classified into 3 main categories: transmission power control, power-aware routing, and use of power save modes.

2.1.1 Transmission Power Control

Power control in mobile ad hoc networks has been the focus of extensive research [4] [5] [6]. Its main objectives are to reduce the total energy consumed in packet delivery and to increase the network throughput by increasing the channel's spatial reuse of the available channels. In this approach, we change the transmission power to adapt to the interference and error rate of the transmission link. Reducing the power reduces the transmission range and decreases the required battery power. In addition, the decreased interference allows greater spatial reuse and this increases the performance of the overall network.

In [17] the authors suggested a protocol that exploits global topological information provided by the routing protocol to reduce the node's transmission power such that the degree of connectivity of each node is upper and lower-bounded. In [18] a cone-based solution that guarantees network connectivity was proposed. The authors in [19] proposed the use of a synchronized global signaling channel to build up global network topology information while each node communicates only with its nearest N neighbors (N is a design parameter). One common deficiency in the above protocols is that they rely solely on Carrier Sense Multiple Access (CSMA) for accessing the wireless channel. It has been shown in [20], [21] that using CSMA alone for accessing the wireless channel significantly degrades network performance.

The ad hoc mode of the IEEE 802.11 standard is by far the most dominant Media Access Control (MAC) protocol for ad hoc networks. This protocol generally follows the paradigm, with extensions to allow for the exchange of Request-To-Send/Clear-To-Send (RTS/CTS) handshake packets between the transmitter and the receiver. These control packets are needed to reserve a transmission period for the subsequent data packets. Nodes transmit their control and data packets at a fixed (maximum) power level, preventing all other potentially interfering nodes from starting their own transmissions. Any node that hears the RTS or the CTS message defers its transmission until the ongoing transmission is over.

For example, the situation in Figure 1, where node A uses its maximum transmission power to send its packets to node B. For simplicity, we assume the used of omni-directional antennas, so a node's coverage floor is represented by a circle in a two-dimensional (2D space). Nodes C and D hear B's CTS message and therefore wait for transmission A → B to finish before attempt to access the medium. However, both transmissions A → B and C → D could take place at the same time if nodes were able to select their transmission powers appropriately, hence, increasing the network throughout and reducing the per packet energy consumption. However, this dynamic reduction in power is not provided in the standard.

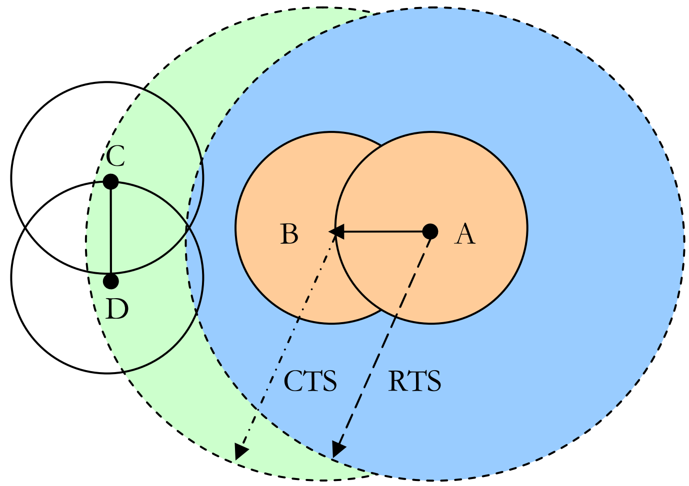

Figure 1: Inefficiency of the standard RTS-CTS approach[1]

The roots of this problem lie in the fact that the IEEE 802.11 standard is based on two non-optimal (in terms of throughput and energy) design decisions:

- An overstated definition of a collision — according to the IEEE 802.11 standard, if node A is currently receiving a packet from node B, then all other nodes in A's transmission range must defer their own transmissions to avoid colliding with A's ongoing reception.

- The IEEE 802.11 standard uses a fixed common transmission power approach, which leads to reduced channel utilization and increased energy consumption.

[1] Nodes A and B are allowed to communicate, but nodes C and D are not. Dashed circles indicate the maximum transmission ranges for nodes A and B, while solid circles indicate the minimum transmission ranges needed for coherent reception at the respective receivers.

From the above example, one can make the following observation: if nodes send their control (RTS-CTS) packets at a fixed maximum power level (P_{max}), but send their data packets at an adjustable (lower) power level, then the collision in the previous example could be avoided. There is no clear indication that reducing the transmit power actually reduces the devices power consumption. In fact, for the early Lucent/NCR WaveLAN cards the reduction in transmission power occurs through attenuation - hence there is no reduction in the device's power consumption.

2.1.2 Power-Aware Routing

Energy saving could be achieved by routing packets over an energy-efficient path [2]**Error! Reference source not found.**[8][10][11][13]. Most of the current routing algorithms are based on the shortest path metric. However, the shortest path does not guaranty the optimal power consumption. Communication between two nodes far way could cost more energy than using multi-hop communication via intermediate nodes. That is because, long range transmission need more power to transmit signal and also lowers the receiver sensitivity which could lead to the overhear problem.

In contrast to conventional wired routing protocols which try to utilize the minimum-hop route, power-aware routing protocols usually aim to utilize the most energy-efficient route. These protocols exploit the fact that the transmission power required on a wireless link is a non-linear function of the link distance, and assume that each node can adapt their transmission power levels. As a consequence of this, it turns out that choosing a route with a large number of short-distance hops often consume significantly less energy than an alternative one with a few long-distance hops [11][43] (The radios all used an identical transmission power independent of the link distance, and if all the wireless links are error-free, then conventional minimum-hop routing).

In practical wireless networks with non-negligible link loss rates, packet retransmission or forward-error correction codes are employed to ensure reliable end-to-end delivery over the entire wireless path. For reliable energy-efficient communication, the routing algorithm must consider not only the distance of each link but its quality (in terms of its error rate) as well. Intuitively, experiments in [3] showed that the cost of choosing a particular link is defined not simply in terms of the basic transmission power but also the overall transmission energy (including possible retransmissions) needed to ensure eventual error-free delivery. This is especially important in practical multi-hop wireless environments, where packet loss rates can be as high as 15–25 %.

Besides, presenting the algorithmic modifications needed to compute a minimum-energy path for reliable communication, conventional routing protocols are "proactive" and compute paths for each (source–destination) pair irrespective of whether those paths are needed or used. This requires the periodic exchange or flooding of routing messages, which can itself consume significant energy, especially when the traffic flows are sparsely distributed. To avoid these overheads, a family of "reactive" routing protocols has been proposed specifically for wireless networks. These protocols (e.g., AODV [13] and DSR [15]) compute routes on demand, when they are needed for a specific traffic flow. Using AODV as a representative protocol, we shall explain the enhancements needed to compute minimum-energy reliable paths with a reactive protocol.

Research about power-aware routing in [1] [7] [8] show that such a routing algorithm reduces the cost/packet of routing packets by 5-30% over shortest-hop routing and even reduces the energy consumption by 50-70% when using protocols such as PAMAS [11] (Power Aware Multi-Access Protocol) or PARO (Power-Aware Route Optimization) [12] . PAMAS is an energy-aware MAC/routing protocol, which proposes to set the link cost equal to the transmission power; the minimum-cost path is then equivalent to the one that uses the smallest cumulative energy. In the variable-power case, where nodes adjust their power on the basis of the link distance, such a formulation often selects a path with a large number of hops. This approach uses a modified form of the Bellman–Ford algorithm. Therefore the selected paths have a smaller number of hops than in the power-aware multi-access protocol with signaling. The power-aware route optimization (PARO) algorithm [12] has also been proposed as a distributed route computation technique for variable-power scenarios, and aims to generate a path with a larger number of short-distance hops. According to the PARO protocol, a candidate intermediary node monitors an ongoing direct communication between two nodes and evaluates the potential for power savings by inserting itself in the forwarding path - in effect, replacing the direct hop between the two nodes by two smaller hops through itself [13].

Alternative metrics, besides the minimum cumulative transmission energy, have also been considered for selection of energy-efficient routes in wireless environments. Indeed, selecting minimum-energy paths can sometimes unfairly penalize a subset of the nodes; for instance, if several minimum-energy routes have a common node in the path, the battery of that node will be exhausted quickly. Researchers have thus used an alternate objective function - maximizing the network lifetime - that considers both the energy consumption of a particular path and the remaining battery capacity of nodes on that path. The key idea is to distribute the energy expenditure across all the constituent nodes, selecting a less energy-efficient path if it helps extend the lifetime of a node nearing battery exhaustion. For example, Singh et al. [2] uses node “capacity” as a routing metric, where the capacity of each node was a decreasing function of the residual battery capacity. A minimum-cost path selection algorithm then helps to navigate routes away from paths where many of the intermediate nodes are facing battery exhaustion. Similarly, the MMBCR and CMMBCR algorithms [8] use a MAX-MIN route selection strategy, choosing a path that has the largest capacity value for its most critical ("bottleneck") node, where the bottleneck node for any given path is the one that has the least residual battery capacity.

2.1.3 Power Saving Modes

There has been extensive research about this topic. The researches in [5][19] show that 802.11 network interface cards consume significant amounts of energy and drain batteries fast, especially in smaller handheld devices. To prolong battery life, the 802.11 standard defines an optional "power-save mode[2]". End users can activate power-save mode via the radio card's vendor-supplied configuration tool (client utilities) or operating system interface. With power-save mode disabled, the 802.11 network card is generally in receive mode listening for packets and occasionally in transmit mode when sending packets. These modes require the client station to keep most circuits powered-up and ready for operation. The important point is how long a node should be put in sleep mode.

2.1.3.1 General Operation

Stations that have their client utilities configured for power-save mode will send all of their frames to the access point with the power management bit in the frame control field of each 802.11 MAC frame header set to 1. This indicates the station's desire to remain in power-save mode, and it informs the access point that it should buffer unicast data frames for the station until polled by the station. This continues to be the case until such time that the station's client utility is

---

[2] For example, the Cisco Aironet 350 Series Client Adapter consumes 2.25 W and 1.35 W in transmit and receive modes, respectively, but consumes only 0.075 W in sleep mode.

reconfigured for fully awake mode. At this time, the station will send its frames to the access point with the power management bit set to 0 to indicate that it is fully awake and the access point should not buffer frames on its behalf.

When dozing, the station consumes much less power than normal by shutting off power to nearly everything except for a timing circuit. This enables the station to consume very little power and still be able to wake up periodically (at a predetermined time) to receive regular beacon transmissions coming from the access point. Each beacon frame contains a Traffic Indication Map (TIM) that identifies which dozing stations have unicast frames buffered at the access point. These buffered frames are awaiting delivery to their respective destinations. The dozing station will wake up to view the TIM in the first beacon it hears. A station may doze at its leisure once in power-save mode. When the station discovers the frames are buffered at the access point, then the station will send PS-Poll frames to the access point until the access point's buffer is empty. Upon receiving a PS-Poll frame, the access point may respond with a single queued data frame or it may send an ACK frame. If the access point responds with an ACK frame, it may then send the queued data frame at its leisure. Each queued data frame is sent in response to an additional PS-Poll frame from the station. As long as there are more queued data frames at the access point, each data frame sent to the station will have the More Data bit (Figure 2) in the Frame Control field of the MAC header set to 1. The last queued data frame will have a More Data bit of 0.

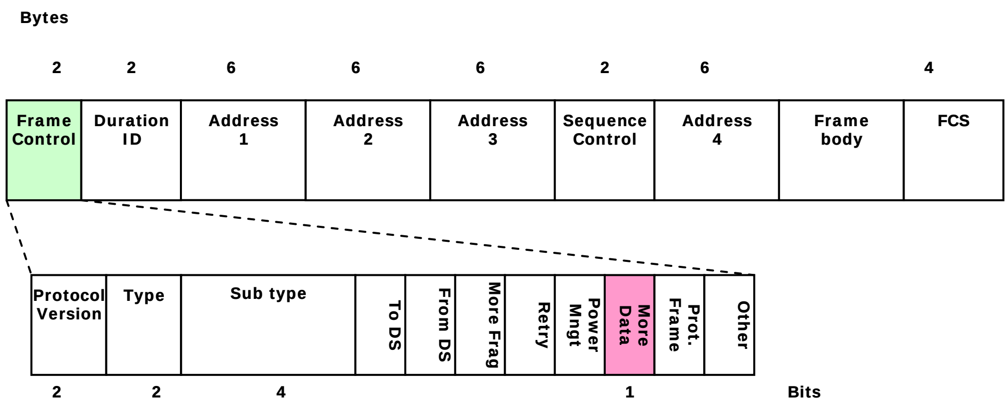

Figure 2: More Data bit

The More Data bit is the method that the 802.11 standard specifies to ensure that stations empty the access point's buffer before dozing again. After emptying the access point's buffer using the PS-Poll mechanism, the beacon will no longer show that station's AID in the TIM. The station may return to "doze" mode at its convenience.

2.1.3.2 Power Management Bit Flipping

The 802.11 standard method of performing queuing and retrieval of data frames at the access point for the benefit of Power-Save stations is not the only method used. The standard calls for use of PS-Poll frames while stations maintain their Power-Save mode in the BSS. Some of the new chipsets on the market perform the same function by flipping the Power Management bit in the Frame Control field of the MAC header (see Figure 2 above) on and off as needed in order to accomplish the same thing as those stations using PS-Poll frames.

This alternate method of queued data retrieval operates as follows:

1. The station sends a Null Function Data frame to the access point with the Power Management bit set to 1. This setting indicates to the access point that the station is going to power-save mode.

2. The access point starts queuing data frames for the station.

3. The station changes to the doze state then periodically powers up and sends a Null Function Data frame to the access point with the Power Management bit set to 0. The station sends this frame without regard to what it might have heard in the beacon.

4. The access point stops queuing data for the station, and immediately sends any queued frames to the station as fast as possible. If there are no queued frames for this station, nothing happens.

5. The station sends a Null Function Data frame with the Power Management bit set to 1 to the access point indicating that the station is returning to power-save mode.

The 802.11 standard states that the More Data field is used to indicate to a station in power-save mode that more MAC Service Data Units (MSDUs) or MAC Management Protocol Data Units (MMPDUs) are buffered for that station at the access point [32]. The More Data field is valid in directed data or management type frames transmitted by an access point to a station in power-save mode. A value of 1 indicates that at least one additional buffered MSDU or MMPDU is present for the same station. The standard does not specify that the More Data bit is used to indicate additional buffered frames for stations that are not in Power-Save mode. This functionality is not expressly specified in the standard, so the decision on when to return to doze state is up to the implementer.

2.1.3.3 DTIMs

When operating in accordance with the 802.11 standard's power-saving mode, the stations must know when to wake up from dozing. Stations using power-save mode will awaken periodically, based on a number of beacons, in order to receive the beacons and to watch for their own AID in the beacon's TIMs. There are two types of TIMs with which a wireless network analyst should be familiar: TIMs and DTIMs. We have discussed that TIMs are used to notify power-save mode stations that they have unicast traffic queued at the access point. This section will discuss DTIMs.

A DTIM is a TIM with particular settings in its fields used to indicate (to the BSS) the presence of queued broadcast/multicast traffic at the access point.

Element ID	Length	DTIM Count	DTIM Period	Bitmap control	Partial Virtual Bitmap
1	1	1	1	1	1-251 octets

Figure 3: TIM Information Element

The DTIM Count Field indicates how many more beacons (including the current frame) appear before the next DTIM. A DTIM count of 0 indicates that the current TIM is a DTIM. The DTIM Period field indicates the number of beacon intervals between successive DTIMs. If all TIMs are DTIMs, the DTIM Period field has the value 1. The DTIM Period value 0 is reserved. The Bitmap Control field is a single octet. Bit 0 of the Bitmap Control field contains the Traffic Indicator bit associated with Association ID 0. This bit is set to 1 in TIM elements with a value of 0 in the DTIM Count Field when one or more broadcast or multicast frames are buffered at the access point. The remaining 7 bits of the

Bitmap Control field form the Bitmap Offset. Each bit in the Virtual Bitmap corresponds to traffic buffered for a specific station within the BSS that the access point is prepared to deliver at the time the beacon frame is transmitted.

The DTIM interval is the interval between TIMs that are DTIMs and is configurable in the access point (or wireless LAN switch). Stations do not request broadcast/multicast traffic, but rather the traffic is delivered automatically following beacons that contain DTIMs. The bit for AID 0 (zero) is set to 1 whenever broadcast or multicast traffic is buffered.

The More Data field of each broadcast/multicast frame is set to indicate the presence of further buffered broadcast/multicast data frames. If the access point is unable to transmit all of the buffered broadcast/multicast data frames before the TBTT following the DTIM, the access point will indicate that it will continue to deliver the broadcast/multicast data frames by setting the bit for AID 0 of the TIM element of every beacon frame, until all buffered broadcast/multicast frames have been transmitted.

2.1.3.4 Ad Hoc

As with infrastructure networks, Ad Hoc stations indicate that they are entering power-save mode by setting the power management bit to 1 in all of their frames. Other Stations have power management bit set to 1 in the IBSS may not transmit data to this station at will, but have to buffer frames locally, send ATIM frames, and then send data frames at appropriate times.

Regularly, all dozing stations wake up at the same time for what is called the Announcement Traffic Indication Message (ATIM) window, which corresponds with each beacon transmission. Stations can choose among many approaches aim to synchronize the state changes in the network through. Distributed beacon generation and introduce mechanisms where nodes synchronously wake up at designated points of time to exchange announcements about pending traffic. Synchronization however is difficult to achieve, in particular in ad hoc networks where all nodes ideally wake up at the same time, at the beginning of a beacon interval, and remain awake during the ATIM window to exchange traffic announcements in case of waiting traffic, and fall back to asleep again if there is no traffic to transfer [41]. If a station is holding frames for a station operating in power-save mode, the station will send an ATIM frame to the power-save mode station indicating that frames are awaiting transmission. The power-save mode station that typically spends its time dozing then knows to stay awake through the next beacon interval, which is hopefully long enough for the station buffering the frame to send the frame successfully. After receiving and acknowledging receipt of the frame, the station can return to a doze state and need not to periodically awake up again to deal with pending traffic.

ATIM frames are messages that contain no frame body. Receiving stations know what ATIM frames are by frame type and subtype. Unicast ATIM frames are acknowledged, but broadcast/multicast ATIM frames are not acknowledged. The ATIM window's length is specified in the beacon's IBSS Parameter Set information element and is measured in Time Units (TUs). 16-bit Beacon Interval value is the number of TUs between Target Beacon Transmission Times (TBTTs). Each TU equals 1024 microseconds (1.024 milliseconds), which are what most vendors referring to as a Kilo-microsecond (Kμs).

The actual savings in battery life using 802.11 power-save modes is difficult to determine, and there are situations where power-save mode might not provide any benefit at all. When transmitting or receiving, the client station will consume an average of 250 milliamps, whereas current draw while dozing could be as low as 30 milliamps at the same voltage level. Because the dozing station will wake up periodically, the aggregate current draw will vary somewhere between 30 and 250 milliamps, depending on the listen interval and doze policy set in the clients.

If the client stays awake longer to accommodate higher traffic levels, then the aggregate current will be closer to the receive/transmit values, possibly 230 milliamps, or so. As a result, the savings in battery life will not be appreciable. Also, keep in mind that to achieve significant battery savings using power-save mode, lower throughput will likely prevail for the power-save stations. In fact, some applications that require frequent communications with the clients will not operate well with power-save mode enabled.

2.2 Issue in 802.11 networks

2.2.1 Protection Mechanisms

The 802.11g amendment to 802.11-1999 (R2003) clearly states that access points (APs) should signal to all associated stations in the basic service set (BSS) to use protection mechanisms (RTS or CTS-to-self) when a NonERP[3] (802.1b) station (STA) associates to the AP.

[3] ERP - Extended Rate Physical (clause 19 802.11 Standard [32]). This clause specifies further rate extension of the PHY (physical layer specification) for the Direct Sequence Spread Spectrum (DSSS) system of Clause 15 and the extensions of Clause 18 (HR-DSSS). This PHY operates in the 2.4 GHz ISM band and builds on the payload data rates of 1 and 2 Mbps, as described in Clause 15, that use DSSS modulation and builds on the payload data rates of 1, 2, 5.5, and 11 Mbps, as described in Clause 18, that use DSSS, CCK, and optional PBCC modulations. ERP-OFDM draws from Clause 17 (OFDM) to provide additional payload data rates of 6, 9, 12, 18, 24, 36, 48, and 54 Mbps. Of these rates, transmission and reception capability for 1, 2, 5.5, 11, 6, 12, and 24 Mbps data rates is mandatory.

IEEE 802.11g, Section 7.3.2.13 states "If one or more NonERP STAs are associated in the BSS the Use_Protection (Figure 4) bit shall be set to 1 in transmitted ERP Information Element[4].

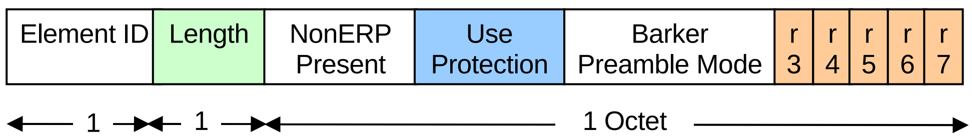

Figure 4: ERP Information Element Format

Using this protection mechanism can easily cause more than a 50% loss in overall WLAN throughput in the BSS. Latency is also increased significantly, more so with RTS/CTS than with CTS-to-Self[5]. The same section of the 802.11g amendment also states "The NonERP_Present bit shall be set to 1 when a NonERP STA is associated with BSS. Example of when the NonERP_Present bit may additionally be set to 1 includes, but is not limited to, when:

[4] The ERP Information element contains information on the presence of Clause 15 (802.11 DSSS) or Clause 18 (802.11b DSSS) stations in the BSS that are not capable of Clause 19 ERP-OFDM (802.11g) data rates. It also contains the requirement that the ERP Information element sender (access point in a BSS or station in an IBSS) is use protection mechanisms to optimize BSS performance and can use long or short Barker preambles. Figure 4 shows the format of the ERP Information Element. If one or more NonERP (802.11 DSSS or 802.11b DSSS) stations are associated in the BSS, the Use_Protection bit should be set to 1 in transmitted ERP Information Elements.

[5] An optional mechanism is designed to guide NonERP stations that a transmission is pending, so that those stations will properly update their NAVs and not transmit during an ERP-OFDM transmission. This mechanism allows ERP stations to exchange frames using the ERP-OFDM modulation that is undetectable by the DSSS or HR/DSSS stations. In a small BSS, without the present of hidden nodes, this mechanism alert NonERP stations to defer for a frame exchange sequence even though the data frame will be undetectable to the NonERP stations. CTS-to-Self is a standard CTS frame transmitted using a NonERP modulation with a destination address of the transmitting station. Obviously the transmitting station cannot hear its own transmission in a half-duplex medium, so the transmission of this frame likes the human vocal equivalent of shouting "Be Quiet!" All nearby stations are alerted that a frame exchange process is pending.

a) A NonERP infrastructure or independent BSS is overlapping

b) In an IBSS, if a Beacon frame is received from one of the IBSS participants where the supported rate set contains only basic rates.

c) A management frames is received where the supported rate set includes only basic rates.

This means if a STA or AP hears a Beacon that has a supported rate set of 11, 5.5, 2, and 1 Mbps (208.11b) or only 2 and 1Mbps(802.11) sent by a nearby AP or a STA that is part of an IBSS, it may enable the NonERP_Present bit in its own Beacons.

Both RTS/CTS and CTS-to-Self have detrimental impacts on throughput. RTS/CTS has greater negative impact on throughput, but a more positive impact on hidden nodes in most wireless environments. It is typical to see that half of a BSS's throughput is lost due to protection mechanisms alone (when they are enabled). Additionally, one ERP access point's decision to enable protection may affect the entire wireless LAN infrastructure as a whole in a negative fashion due to vendor-specific implementations.

Each ERP access point's beacon has an ERP Information Element. The Barker_Preamble_Mode bit is used to specify whether long or short preambles are to be used when transmitting frames modulated with BPSK, QPSK, or CCK. These frames include RTS, CTS, and data fragments (except the last fragment) that are transmitted by an ERP station that has been notified by the access point, using the NonERP_Present field, that a NonERP station is associated with the access point. The Use_Protection bit is used by the access point in beacons to alert ERP stations in the BSS that they should use protection mechanisms such as RTS/CTS and CTS-to-Self before transmitting data using OFDM modulation.

Depending on the vendor's implementation of the standard, co-channel and adjacent channel interference between access points may also lead to a situation in which an access point will enable protection in its own beacons if it hears another access point enable protection in its beacons. This reaction can lead to a situation in which a single NonERP station causes protection mechanisms to be enabled throughout part or all of a wireless LAN infrastructure.

Suppose that a NonERP station successfully roams to another access point in the ESS. This roaming will cause the new access point to enable protection and notify the old access point this particular mobile station has now re-associated with this access point. This notification should cause the old access point to drop the association with the NonERP station immediately. When the new access point enables protection, its beacons may then cause the old access point to either keep protection enabled or to immediately re-enable it. The old access point may make this change even though the NonERP station has left its BSS.

This cause and effect scenario demonstrates that everywhere a NonERP station goes in an ERP WLAN, protection mechanisms are not only enabled on the local access point, but may also be triggered elsewhere depending on which access points can hear which other access points. The chance that any single access point can hear at least one other access point is very good in most enterprises. When a NonERP station roams, it causes a wave of "Use_Protection=0" from the old access point immediately followed by a wave of "Use_Protection=1" from the new access point across the wireless LAN as shown in Figure 5. Imagine how many times this scenario takes place when there are many NonERP clients roaming about the enterprise.

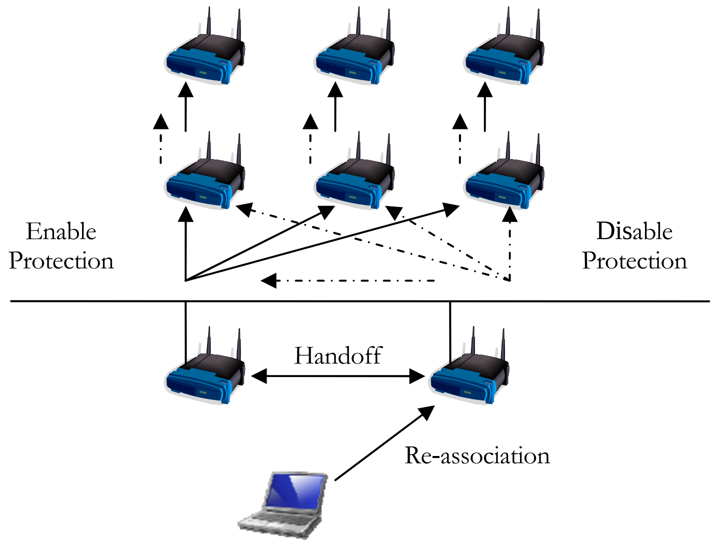

Figure 5: Use_Protection Proliferation

Throughput degradation when using protection in a BSS is severe. The loss in throughput is approximately half for the entire BSS even if the 802.11b station just associates. Thus the 802.11b station does not have to transmit any traffic for this to happen, but doing so makes the situation far worse [42].

2.2.2 RTS/CTS

As an optional feature, the 802.11 standard includes the Request-to-Send/Clear-to-Send (RTS/CTS) function to reserve medium access. With RTS/CTS enabled, a station may transmit a data frame after it completes an RTS/CTS handshake with the immediate receiver of the data frame.

A station (or access point) initiates the four step frame exchange sequence by sending an RTS frame to the intended receiver of the subsequent data frame. The immediate receiver of the RTS responds with a CTS frame.

The station that sent the RTS frame must receive the CTS frame before sending the data frame. The RTS and CTS frames each contain values in their duration fields that indicate the amount of time needed to complete the transfer of the subsequent data frame and acknowledgement. This duration field value alerts nearby stations to hold off from transmitting for the duration of the four step frame exchange sequence.

The RTS/CTS protocol provides positive control over the use of the shared medium. The purpose of the RTS/CTS protocol is to reserve the wireless medium in order to minimize collisions among hidden stations. This "hidden node" problem can occur when users and access points are spread out throughout a facility or when 802.1b and 802.11g stations coexist in the same BSS or BSA. Using the RTS/CTS protocol to alleviate collisions in this kind of scenario is a "protection mechanism" as described earlier. The main difference between use of the RTS/CTS protocol as a manually-configured medium reservation tool and use of the RTS/CTS protocol as a protection mechanism is that when RTS/CTS is used as a protection mechanism, it is automatically enabled by the access point's beacons.

Duration Values & Modulation: There are a complex set of rules regarding duration values specified by the 802.11g standard [22] , which will be translated into layman's terms bellow. In a fragment burst (i.e., a series of frames which are due to fragmentation of a frames which exceeds the link MTU and are sent as a burst), the modulation of frames is as follows. NonERP modulation is used with:

- RTS & CTS frames
- All ACK frames except the last ACK frame in the fragment burst
- All data fragments except the last fragment in a burst

ERP modulation is used with:

- The last ACK frame in a fragment burst

- The last data fragment in a burst

ACK frames should be sent at the same rate and modulation as the data frame which preceded it. If they are not, then the station that transmitted the data frame may not understand the ACK and may begin retransmissions.

If a protection mechanism, such as RTS/CTS, is being used, a fragment sequence may only employ ERP-OFDM modulation for the final fragment and control response because the duration values of data fragments and their corresponding NonERP-modulated data fragment and ACK frames are used as a virtual RTS and CTS for subsequent fragments and Acknowledgement frames (ACKs). In order to be understood by NonERP stations, all but the last fragment and ACK must be sent using a modulation that NonERP stations will understand.

Each ACK frame sets the NAV of NonERP (and ERP) stations in the BSS and BSA to a value equal to the subsequent SIFS+DATA+SIFS+ACK (ACKs that are not the last ACK) or to a value of 0 (the last ACK). The data fragments (except the last fragment) also set the NAV of ERP and NonERP stations in the BSS by having a duration value equal to that of the subsequent SIFS+ACK. Therefore, these data fragments must also be sent using a modulation type that is understood by NonERP stations. The last data fragment and ACK are covered by the duration value of the immediately preceding ACK frame, so they can be transmitted using ERP-OFDM without any problems.

Notice in Figure 6 and Figure 7 that each frame contains a duration value equal to subsequent inter-frame spaces and frames in accordance with the rules listed above. A frame's duration value never takes into account its own length, but rather a certain number of inter-frame spaces and frames that come after it in a frame exchange sequence. In Figure 6, the data frame and ACK may be transmitted using ERP-OFDM modulation because there is no fragmentation in use.

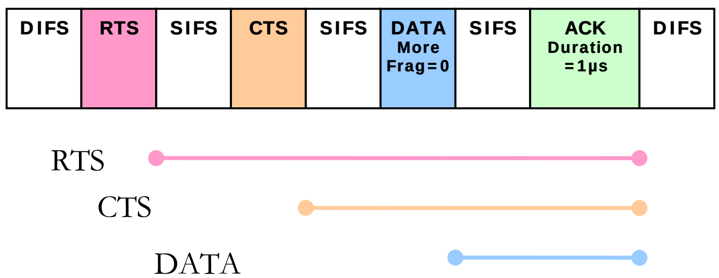

Figure 6: Duration – no Fragmentation

RTS and CTS frame duration values only provide for the first data fragment and its corresponding acknowledgement as shown in Figure 7. The duration value found in subsequent data fragments and their ACK frames reserve the medium for enough time for the next fragment and ACK. In Figure 7, the first data fragment must use NonERP modulation (such as BPSK, QPSK, or CCK), and the second data fragment will use OFDM modulation. There is no need for the last data fragment and ACK to be understood (for NAV-setting purposes) by NonERP stations in the BSS because the previous ACK reserved the medium using NonERP modulation.

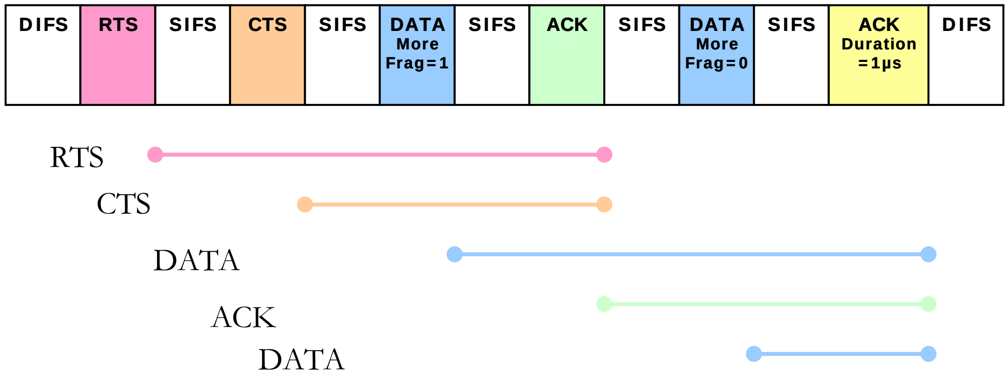

Figure 7: Duration - Fragmentation

RTS/CTS can be effectively disabled by setting the threshold value to the highest available value as shown in the client utilities screenshot in Figure 8 (note that this utility is for the Windows 2000/XP operating system).

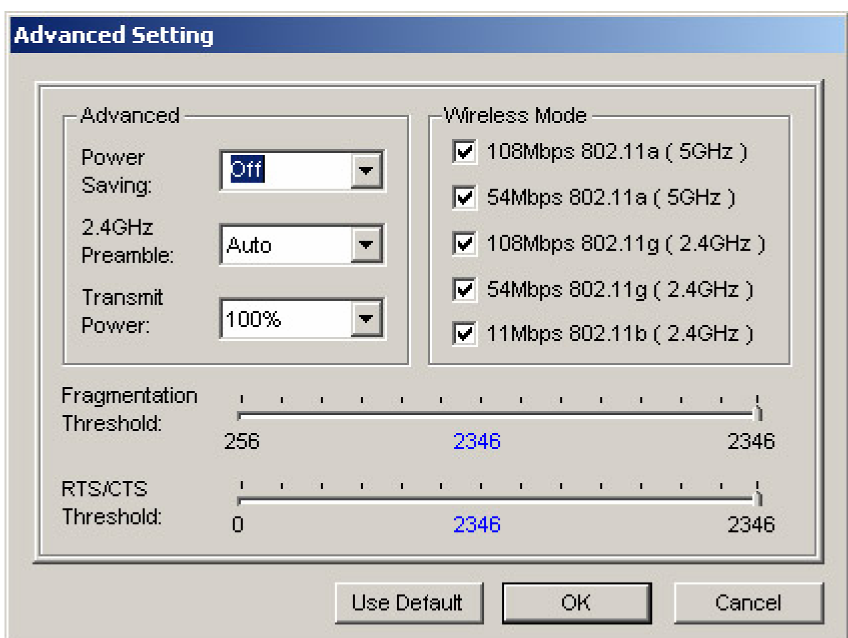

Figure 8: Configuring the RTS/CTS Threshold

Some vendors provide "on" and "off" software settings in addition to the threshold value only. RTS/CTS can be enabled all the time by setting the threshold value to the lowest available value of 0. Keep in mind that an increase in performance using RTS/CTS is the net result of introducing overhead (i.e., RTS/CTS frames) and reducing overhead (i.e., fewer retransmissions). If you do not have any hidden nodes, then the use of RTS/CTS will only increase the amount of overhead, which reduces throughput. A slight hidden node problem may also result in performance degradation if you implement RTS/CTS. In this case, the additional RTS/CTS frames cost more in terms of overhead than what you gain by reducing retransmissions.

As with fragmentation, one of the best ways to determine if you should activate RTS/CTS is to monitor the wireless LAN for retransmissions. If you find a large number of retransmissions and the users are relatively far apart and likely out of range, then try enabling RTS/CTS on the applicable user wireless NICs. Keep in mind that user mobility can change the results. A highly mobile user may be hidden for a short period of time, perhaps when you perform the testing, then be closer to other stations most of the time. If collisions are occurring between users within range of each other, the problem may be the result of high network utilization or possibly RF interference. In this case, RTS/CTS can be turned off.

Because RTS/CTS introduces overhead, you should shut it off if you find a drop in throughput, even if you have fewer retransmissions. After all, the goal is generally to improve performance. Except in the case of access points contending for the same channel in the same BSA, initiating RTS/CTS in the access point is not useful because the hidden station problem does not exist from the perspective of the access point. All stations having valid associations are within range and not hidden from the access point. Forcing the access point to implement the RTS/CTS handshake will significantly increase the overhead and reduce throughput.

2.3 Internet connectivity for mobile ad hoc networks

In spite of the fact that, a MANET is useful in many situations such as emergency, battle field, disasters, or in remote area, the ability to connect to the Internet is generally highly desirable. This internetworking is achieved by using gateways, which act as bridges between a MANET and the Internet. In order to communicate with a host located on the Internet a mobile node in the MANET needs to find a route to a gateway. This requires gateway discovery.

The ad hoc routing protocols were designed for communication **within** a MANET. Therefore, the routing protocol needs to be modified in order to provide bridging capability between a mobile device in a MANET and a fixed device in a wired network. To achieve this network interconnection, gateways that understand the protocols of both the MANET protocol stack and the TCP/IP suite are needed. All communication between the two networks must then pass through the gateway. Gateways expand the communication beyond an ad hoc network, but require some last hop mobility management.

Two classes of approaches have been proposed to support connectivity between ad hoc networks and the Internet.

- Proactive schemes flood advertisements from nodes through the whole ad hoc network to find the gateway. Such approaches provide good connectivity, but impose a high overhead, especially when not all the nodes in the ad hoc network require external connectivity.

- Reactive schemes allow the mobile nodes to broadcast solicitations to find nodes and gateways as they are needed. Such approaches keep the overhead of maintaining connectivity to external networks low, but negatively impact on the mechanisms necessary for gateway discovery and movement detection.

- Hybrid scheme that combines proactive and reactive techniques to provide connectivity with reduced overhead. In our approach, gateway discovery advertisements are flooded within a limited number of hops. Nodes that are outside this hop limit use reactive techniques to solicit foreign agents when needed. A hybrid approach combines the advantages of both proactive and reactive approaches and provides good connectivity while keeping overhead costs low.

Choosing an addressing scheme is also an important issue when designing gateway discovery protocol for MANET. Two popular approaches are: Mobile IP and IPv6. Mobile IP using the traditional IPv4 addressing scheme and TCP/IP protocol stack is easy to deploy. However, mobile IP requires additional mechanism to handle problems of triangle routing, keep session alive when roaming… IPv6 solves the scalability problem and provide a unified architecture, but nodes in both wired and wireless domain need to change addressing architecture in order to communicate with each other. Here we will use the IPv6 solution which provide a better scalability and a complete solution [36].

2.3.1 Proactive gateway discovery

The proactive gateway discovery is started by the gateway itself. The gateway periodically broadcasts the Gateway Advertisement messages which are transmitted after expiration of the gateway's timer (ADVERTISMENT_INTERVAL). The time between two consecutive advertisements must be chosen with care so that the network is not flooded unnecessary often. All mobile nodes residing in the gateway's transmission range will receive the advertisement.

When the advertisement is received, the mobile nodes that do not have a route to the gateway create a route entry for it in their routing table. Mobile nodes that already have, update the entry for it. Next the advertisement is forwarded by the mobile nodes to other mobile nodes residing within their transmission range. To assure that all mobile nodes within the mobile ad hoc network receive the advertisement, the number of transmissions is determined by NET_DIAMETER defined by the protocol. However, this will lead to unnecessary duplicated advertisements and this is a disadvantage of this mechanism. However, we can solve this problem by comparing the RREQ ID with the original IP address.

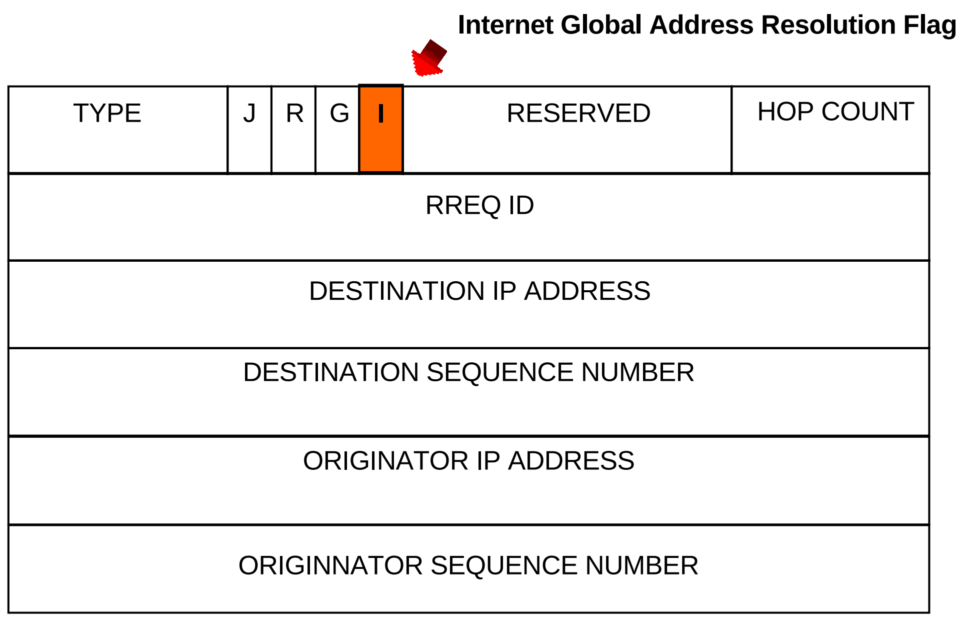

Figure 9: RREQ_I message format

An advertisement is approximately a RREP_I message and since this message does not contain any field similar to the RREQ ID field in RREQ messages, a new AODV message has been introduced: Gateway Advertisement (GWADV). This message is basically a RREP message extended with one field from the

RREQ message, namely the RREQ ID field. When a mobile node receive a GWADV, it first checks to determine whether a GWADV with the same originator IP address and RREQ ID already has been received during the last BCAST_ID_SAVE seconds. If such a GWADV message has not been received, the message is rebroadcasted (after decrementing the life time). Otherwise, the newly received GWADV will be discarded. Hence, duplicate GWADVs are not forwarded and the advertisement is flooded through the whole network without causing too much congestion. However, the disadvantage of this solution is the fact that a new AODV message is introduced which requires AODV to be modified.

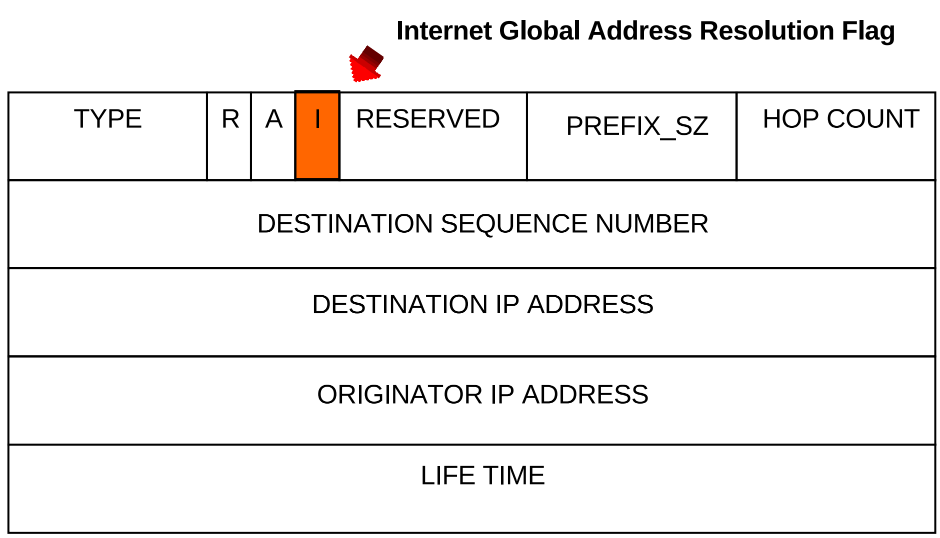

Figure 10: RREP_I message format

2.3.2 Reactive gateway discovery

Unlike the previous mechanism, reactive gateway discovery is initiated by a mobile node that wants to find or update information about a gateway. The mobile node broadcasts a RREQ_I (I stands for Internet Global Address

Resolution flag, this is an extension to the standard RREQ message) to all members of its multicast group. Thus, only the gateways are addressed by this message and only they will process it. Intermediate nodes that receive the message simply forward it by broadcasting it again after decrementing the time to live. When received a RREQ_I, a gateway unicasts back a RREP_I containing the IP address of the gateway [30][28].

The advantage of this approach is that RREQ_I is only sent when mobile node needs information about the reachable gateways. Hence, periodic flooding of the complete mobile node ad hoc network, which has obvious disadvantages, is eliminated. The disadvantage of reactive gateway discovery is that the load on forwarding mobile nodes, especially on those close to a gateway, is increased.

2.3.3 Hybrid gateway discovery

To minimize the disadvantages of proactive and reactive gateway discovery, the two approaches can be combined. This results in a hybrid method for gateway discovery. For mobile nodes within a certain range around a gateway, proactive gateway discovery is used. Mobile nodes residing outside this range use reactive gateway discovery to obtain information about the gateway.

The gateway periodically broadcasts a RREP_I message which is transmitted after expiration of the gateway's timer (ADVERTISEMENT_INTERVAL). All mobile nodes residing in the gateway's transmission range receive the RREP_I. Upon receipt of the message, the mobile nodes that do not have a route to the gateway create a route entry for it in their routing tables. Mobile nodes that already have a route to the gateway update their entry for it. Next, the RREP_I is forwarded by these mobile nodes to other mobile nodes residing in their transmission range. The maximal number of hops a RREP_I can move through the ad hoc network is the ADVERTISEMENT_ZONE.

When a mobile node residing outside this range needs gateway information, it broadcasts a RREQ_I to the ALL_MANET_GW_MULTICAST address. Mobile nodes receiving the RREQ_I simply rebroadcast it. Upon receipt of this RREQ_I, the gateway unicasts back a RREP_I.

2.4 Energy consumption model for Internet connectivity in MANET

The more closely a simulation reflects specific hardware, the more accurate its estimate of the energy consumed. The energy consumption model and simulation environment were chosen to balance these goals: a precise estimate of energy consumption and high-level insight into protocol behavior. The CMU Monarch Project's mobility-enhanced ns-2 simulation environment models the IEEE 802.11 MAC layer, logging control and data messages. The energy consumption model was therefore built based-on the IEEE 802.11 protocol, rather than electronic properties such as mode switching and signal response. Experimental results reflecting the observed energy consumption of an IEEE 802.11 wireless interface were incorporated into the model, providing a quantitative example of energy consumption [23].

The network interface has four possible energy consumption states: transmit and receive are for transmitting and receiving data. In the idle mode, the interface can transmit or receive. This is the default mode for a node in an ad hoc environment. The sleep mode has extremely low power consumption. The interface can neither transmit nor receive until it is woken up. A base station moderates communication among mobile nodes, scheduling and buffering traffic so that the mobiles can spend most of their time in the sleep state.

In an ad hoc environment, there are no base stations and therefore nodes cannot predict when they will receive traffic. The default state of a node in ad hoc networks is idle. The model assumes that the same link-layer operation always has the same costs: an assumption that may not be true if, for example, signal strength affects the energy required to receive the data.

Inconveniently, wireless network interface card (NIC) specifications do not provide information about power consumption in these different modes. Due to the existing indirect nature of the measurements, these values have considerable uncertainty (as much as 5– 10%). Nevertheless, they provide a good indication of relative costs, which is most important for high level analysis. In [23], the study about a detailed of an energy consumption model also gives some keys property which were used in the model used in this thesis:

- The cost of receiving is significant because if a broadcast message is received by more than about four neighbors, the total cost of receiving the packet is greater than the cost of sending it. The relative cost of receiving is likely to increase, reflecting a trend toward greater sensitivity and signal processing capabilities at the receiver.

- The fixed cost of sending or receiving a packet is relatively large compared to the incremental cost. For small packets (130 bytes broadcast or 230 bytes point-to-point), the fixed cost is greater than the incremental cost of sending or receiving a byte. This implies, for example, that small ROUTE_REQUEST or "HELLO" messages are a relatively expensive mechanism. It also suggests that source routing headers are relatively inexpensive in terms of energy consumption.

- Discarding a packet is generally much less expensive than receiving it. With large messages, non-destination nodes can reduce their energy consumption while data is being transmitted and therefore significant reduce energy consumed to receive and process the packet if they can quickly determine that the packet is not relevant to them and then enter sleep mode for the duration of the packet..

CHAPTER 3: SIMULATION ENVINRONMENT

3.1 Topology

We are using Ns-2 [34], a highly modular discrete event simulator, developed for simulating the behavior of network and transport layer protocols in a complex network topology. It is freely available and has been extensively enhanced by the Monarch Project at CMU [37] for use in simulating wireless ad hoc networks.

At the physical layer, the radio model supports propagation delay, and a two-ray ground reflection radio propagation model. At the link layer, the IEEE 802.11 MAC protocol and Distributed Coordination Function (DCF) for ad hoc use are supported. The scenarios used in these simulations were designed based on the IMPORTANT (Impact of Mobility Patterns On RouTing in Ad-hoc NeTworks) mobility framework and Internet connectivity scenario [30]. The scenarios reflect relatively dense networks with potentially high levels of node mobility and hence connectivity changes. The traffic load was low bandwidth, but with a fairly high endpoint diversity.

• Transmit and receive characteristics were based on specifications for the LucentWaveLAN 2.4 GHz DSSS IEEE 802.11 PC card, which has a nominal data transmission range of 250 m. Compared to these older WaveLAN cards, newer cards have greater receive sensitivity and nominal transmit range.

• 48 mobile nodes moved around a 1000m × 1000m area for 300 s of simulated time. When there are few nodes in network and mobile nodes want to connect to nodes outside ad hoc network, it needs to send gateway discovery message to almost every nodes in ad hoc network. As the result other node have to stay awake to response to the require nodes or forward intermediate traffic. Early studies on simulation scenarios using 12 and 24 nodes show that the energy consumption of nodes is not much different.

• 12 randomly chosen source-destination pairs provided traffic load. Each source sent a constant bit rate stream consisting of four 64-byte IP packets/s to its destination. In highly dynamic and heavy traffic, nodes in MANET have to always stay awake to carry traffic and therefore energy variation is low. The similar problem about the impact of load is showed in [5].

A node could act as the source or sink for more than one stream and streams were jittered to avoid artificial interactions.

To support Internet connectivity, modifications to the ns-2 simulator were required based on [30]; note that only logging functionality was added. Energy consumption calculations were done entirely via post-processing. Appendix A shows detail of the implementation and installation instructions to build the simulation scenarios and analyze the trace files generated.

It is nontrivial to differentiate the energy consumed on a per-packet basis based upon whether a node is in range of the sender, the destination, or both. There is also some difficulty in establishing a structure in the experimental measurements of the small cost of discarding control packets. Therefore, to compute the cost of discard traffic, all nodes in range of the sender are assumed to be in range of the receiver. This overestimates the cost of discarding packets, as nodes in range of the sender, but not the destination are erroneously charged with energy costs associated with discarding the destination side of the control sequence. Suppose that nodes are equally likely to be in range of the sender, but not the destination, and vice versa, and then the resulting error will be negligible, especially as the cost of discarding control traffic is small. The energy cost of discarding retransmitted control messages is also ignored in all scenarios.

3.2 Mobility and traffic patterns

The experimental environment used IMPORTANT (Impact of Mobility Patterns On RouTing in Ad-hoc NeTworks) framework as its mobility patent generator [33][39]. In this framework, mobility is viewed as a multi-dimensional evaluation space, with each dimension representing a specific mobility characteristic. Various protocol independent metrics are proposed to capture interesting mobility characteristics of a mobility space and connectivity graph. By using a rich set of parameterized mobility models (including Random Waypoint, Random Walk, Reference Point Group Mobility, Freeway, Manhattan, and City Section models), several 'test-suite' scenarios are chosen to carefully span the mobility space. With Freeway, RPGM, and Manhattan mobility patterns, the speed of the nodes are changed from slow (1km/h) to fast (60km/h). The change is translated in terms of network topology structure is changed from static to highly dynamic. Similar parameters were also used for the Random Waypoint scenario where the "pause time" of mobile nodes in the area is changed from 0s to 300s; this has the opposite effect as it changes the network topology from being highly dynamic to essentially static.

3.3 Power saving modes

3.3.1 Energy consumption states

The network interface has four possible energy consumption states: TRANSMIT (TX) and RECEIVE (RX) are for transmitting and receiving data. In the IDLE mode, the interface can transmit or receive. This is the default mode for ad hoc environment. The SLEEP mode has extremely low power consumption. The interface can neither transmit nor receive until it is woken up. A base station moderates communication among mobile nodes, scheduling and buffering traffic so that the mobiles can spend most of their time in the sleep state. In an ad hoc environment, there are no base stations and nodes cannot predict when they will receive traffic. Therefore, the default state in an ad hoc network is the IDLE state, rather than the sleep state. The model computes costs relative to the idle state. As there is currently little work in the area of energy management for ad hoc networks, the model does not provide for arbitrary transition to the SLEEP state. The model assumes that the same link-layer operation always has the same costs: an assumption that may not be true if, for example, signal strength affects the energy required to receive the data.

3.3.2 Energy level setting

The energy level setting is based on propagation model, range of transmission, gain of the antenna, system lost and frequency used. The threshold utility support generates the transmission (TX), received (RX), sleep, idle value and other energy related follow 802.11 standards. Difficulties arise when choosing the suitable energy level of mobile node because of the diversity of the simulation environment. The study in the previous chapter shows that, the protection mechanism will cause the node to use nonERP mechanism instead of ERP. Therefore, the NonERP was chosen to reflect of the actual scenarios.

3.5 Metrics and Parameters

3.5.2 Minimize Energy consumed

Minimizing the energy consumed is one of the most obvious metrics that reflects our intuition about conserving energy. Assume that some packet j travels via nodes $n_1, \ldots, n_k$ where n_1 is the source and n_k the destination. Let $T(a, b)$ denote the energy consumed in transmitting (and receiving) one packet over one hop from u to b. Then the energy consumed for packet j is,

$$e_j = \sum_{i=i}^{k-1} T(n_i, n_{i+1})$$

3.5.3 Minimize Maximum Node Cost

Let $C_i(t)$ denote the cost of sending packet through node i at time t. Define d(t) as the maximum of the $C_i(t)$s. Then,

$$\hat{C}(t), \forall t > 0$$

This metric minimizes minimum node cost. An alternative is to minimize the maximum node cost after routing N packets to their destinations or after T seconds. All of these variations ensure that node failure (due to exhausting its limited power supply) is delayed; a side effect is that the variance in node power level is also reduced. Unfortunately, we see no way of implementing this metric directly in a routing protocol. Moreover, minimizing cost/node does significantly reduce the maximum node cost (and hence time to first node failure). The metrics discussed above do, in different ways, express our intuition about conserving energy in the network by selecting routes carefully. However, what protocols best implement these metric? It is easy to see that any protocol that has shortest paths can be used to determine optimal routes based on the first and fourth metrics discussed above. To implement the first metric, we simply associate an edge weight with each edge in the network.

3.6 Simulation results

3.6.1 Random waypoint model

The Random Waypoint model is the most commonly used mobility model in research community. In the current distribution of ns2, the implementation of this mobility pattern is as follows: at every instant, a node randomly chooses a destination and moves towards it with a velocity chosen uniformly randomly from $[0,V_{max}]$, where V_{max} is the maximum allowable velocity for every mobile node. After reaching the destination, the node stops for a duration defined by the "pause time" parameter. After that, node chooses another random destination and repeats the whole process until the simulation ends. Figure 1Figure 11 shows the simulation in Random waypoint scenario, 48 mobile nodes (green circle) are distributed randomly in the simulated area.

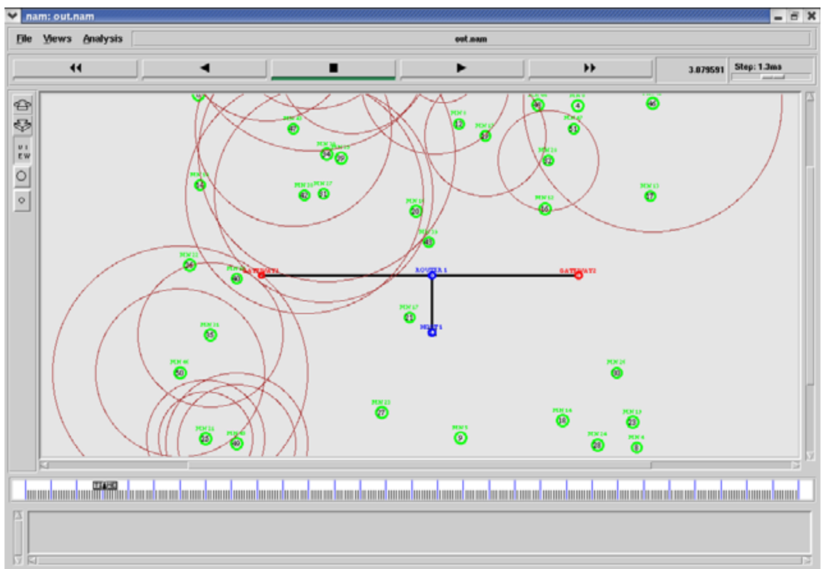

Figure 11: Random waypoint mobility pattern

3.6.2 Freeway mobility model

This model emulates the motion behavior of mobile nodes on a freeway (show in Figure 12). It can be used in exchanging traffic status or tracking a vehicle on a freeway. There are many lanes in the freeway and nodes can move on those lanes.

The speed of node is restricted based on the node moving ahead and a defined safe separation distance.

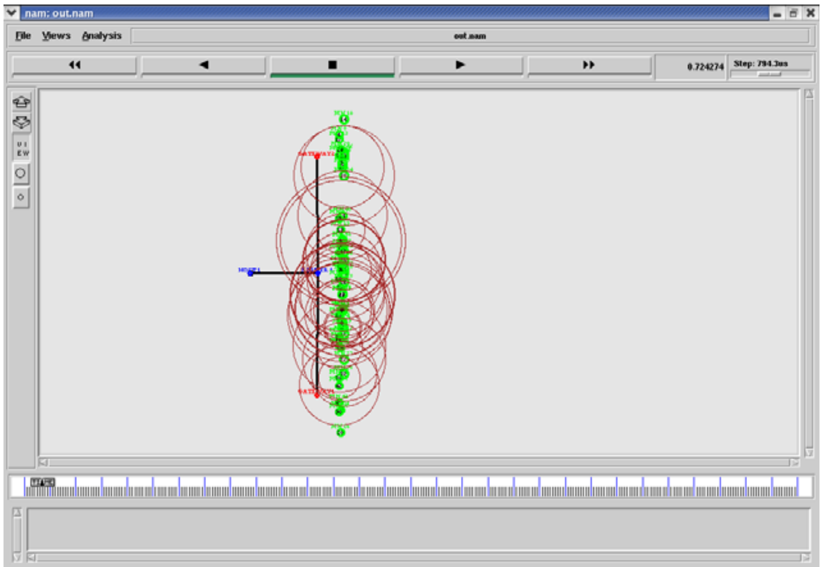

Figure 12: Freeway mobility pattern

3.6.3 Group mobility model

Group mobility [39] can be used in military battlefield and command based communication. Here, each group has a logical group leader. Figure 11 shows that the movements of other nodes are determined by the group leader's behavior; including location, speed, direction, and acceleration. Initially, each member of the group is uniformly distributed in the neighborhood of the group leader. Subsequently, at each instant, every node has a speed and direction that is derived by randomly deviating from that of the group leader.

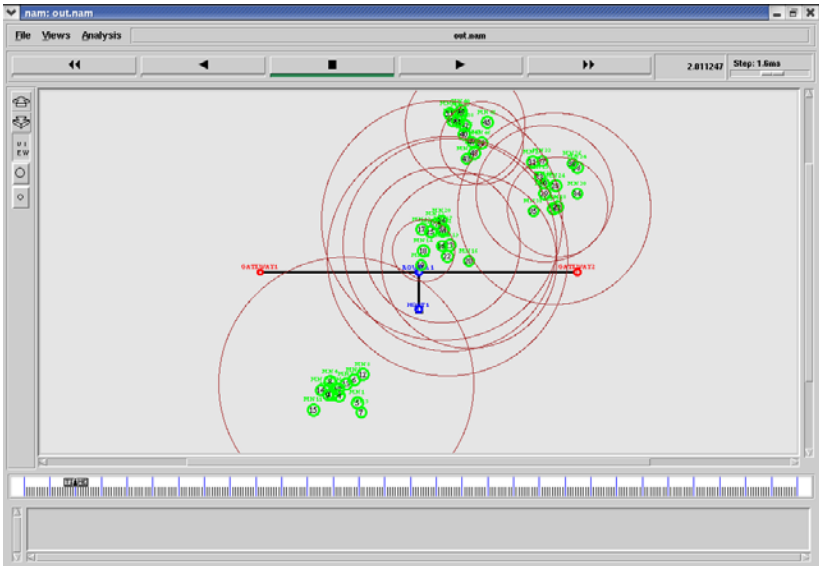

Figure 13: Group Mobility Pattern

3.6.4 Manhattan mobility model

Maps are used in this model. Figure 14 shows 48 mobile nodes moving in the simulation map which is formed by 3 horizontal streets and 3 vertical streets. However, the map is composed of a number of horizontal and vertical streets. The mobile node is only allowed to move along the grid of horizontal and vertical streets on the map. At an intersection of a horizontal and a vertical street, the mobile node can turn left, right, or go straight with a certain probability. Except the above difference, the inter-node and intra-node relationships involved in the Manhattan model are very similar to the Freeway model. Figure 14 show 48 mobile nodes moving in the map-based scenario, there are 3 vertical and 3 horizontal streets in the map.

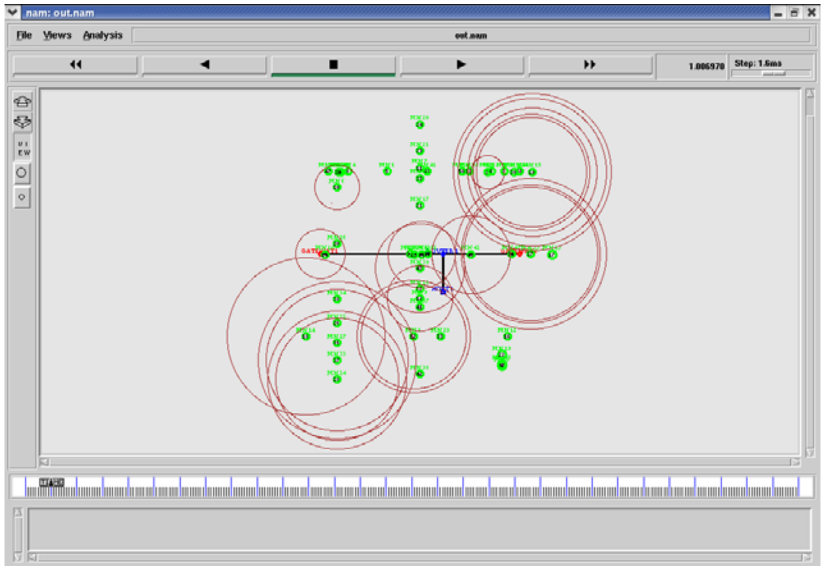

Figure 14: Manhattan Mobility Pattern

CHAPTER 4: ANALYSIS AND DISCUSSION

4.1 Effect of changing the number of gateways

Changing the number of gateways in ad hoc networks not only has a big impact on the performance of the system, but also can make a significant difference in energy consumption of the mobile node. In this simulation, different numbers of gateways are placed in a square area (1000m x 1000m) in order to maximize the network coverage. In practice, the question of where/how to place these gateways is result of a site survey. When deploy wireless network, a site survey provides guidance for the deployment process which includes find out dead-end and maximizing network coverage... In these simulations the placement of the gateways to the Internet were chosen to be uniformly distribute over the square simulated area.

By placing different number of gateways in the simulation area, the results show that the energy consumption of mobile nodes is different under all mobility patterns. By increasing the number of gateways, the residual energy of the nodes is increased significantly. Viewed another way, the total energy consumption decreases because the mobile node can chose an alternative way to reach to the gateway. Increasing the number of gateways makes the routes shorter and therefore decreases the number of route hops required by the packet to reach its destination. The energy consumed when changing the number of gateway from 1 to 2, and from 2 to 3 is quite substantial. The energy utilization ratio is between 15 – 25% of total energy consumed (Figure 15).

If there are only few gateways, in order to reach the nodes outside ad hoc network, the number of hops to reach the gateway point is greater than when there are many numbers of gateways. Although nodes can choose different paths, there is only one default gateway. Furthermore, the number of hops to reach the destination is still high and changing the default gateway requires activating the gateway discovery protocol which will consume a lot of power.

However the energy decrease is not simply counter proportionally to the number of gateways. Once the coverage of gateways reaches a threshold, increasing the number of the gateway does not further reduce the energy used. Sometimes, it will slightly increase the energy consumption level, because it will make the gateway discovery protocol more complicated (about a 5% increase in our simulation).

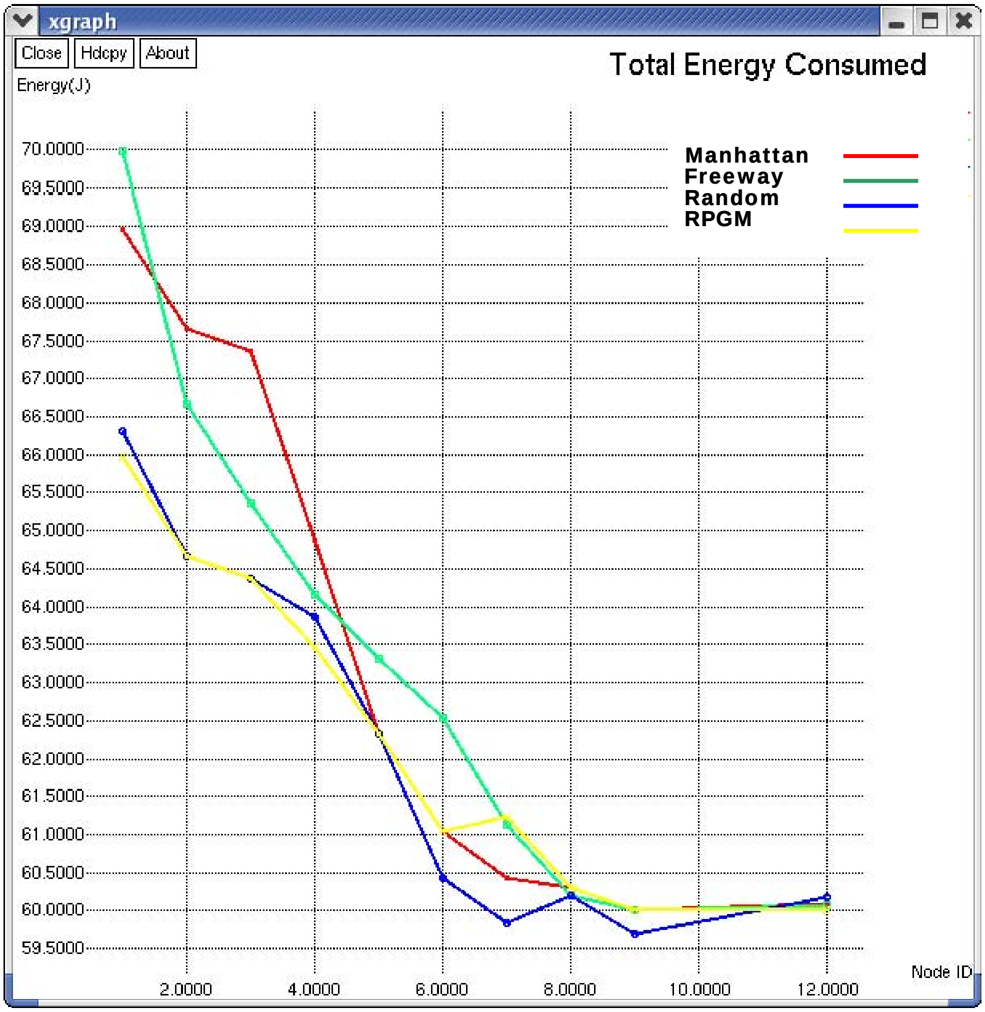

Figure 15: Energy as a function of number of gateways

In a small number of gateways scenarios, the Freeway and Manhattan mobility model consumes more energy than other scenario. The reason is that, in such scenarios, the packet needs to be forwarded through a large number of intermediate nodes before reach the gateways. This increase in the number of forwarding node requires more energy. Our simulation results show that, about 10-15% of extra energy was spent for this forwarding along strings of nodes.

Moreover, in practice, the shortest path sometimes does not mean the optimal path regarding energy consumption. When we added a multi-state error model, the results show that optimal routes also depend on the link state and quality. Packets transmitted on high error rate and high latency links are usually dropped due to the queue being full or due to a timeout; both increase the node's energy consumption significantly (Figure 16) due to retransmission.

In the low mobility scenario and in the cases where there are only a few nodes in the network, the result shows that the energy consumption of mobile node is very similar. In these scenarios, most of the nodes have to be awake all the time to carry forward traffic for other nodes; as there are a few options for nodes to choose to reach to the destination. In the forwarding state, the mobile node's consumption energy is as great as in receive or transmits state; hence for the total energy consumption is high. Another reason is that, the network may be partitioned because a node that is far from the other nodes may not find the way to reach the gateway and therefore is unable to send traffic to its desired destination.

Under high load, total consumed energy consumed in the simulation is also high. The reason is that gateways have to carry traffic from multiple nodes to the Internet; hence they quickly become network bottleneck. This could lead to the situation where the queue at the gateway is full, hence packets will be dropped and therefore they will need to be retransmitted if the source traffic is TCP. Each of these re-transmission cost additional energy.

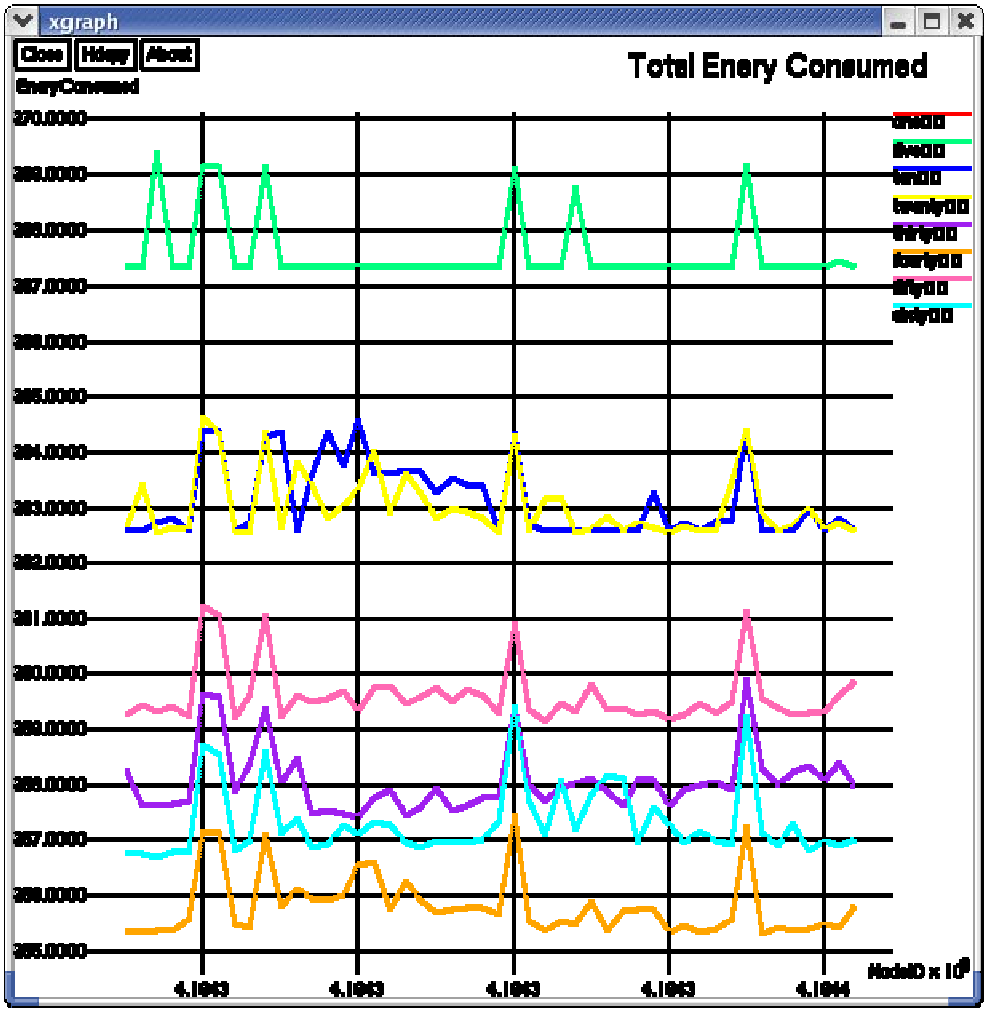

Figure 16: Energy consumed using multi-state error model

Placing gateways in different positions in the network also shows a difference in energy consumption rates. If packets are forward a large number of hops, then extra energy will need to forward these packets.

4.2 Effect of different mobility patterns

Studies show that, Random Waypoint, a common mobility pattern used for simulation and analysis the performance of MANET, does not seem to capture the mobility characteristics of spatial dependence[6], temporal dependence[7] and geographic restriction[8] [33]. At a conference, in a classroom, or in a meeting, mobile nodes are distributed randomly in an area which will exhibit a changing network topology. In a city, the nodes are moving along the definite pathways or streets way where node's movement is restricted by obstacles (often these are shown on maps). On a battlefield, the movement of soldiers is influenced by their commanders. Therefore the movement patterns are different leading to a change in where packets are to be forwarded.

That being said, autonomous architectures are still broadly implemented throughout myriad enterprises and are quite likely to meet the current business needs of the organization that are using them. For new implementations, however, the market place appears to be shifting quite strongly towards different architectures. A depth study of realistic models of the motion patterns required for successfully development and deployment of high performance cost-effective solutions.

---

[6] Spatial dependence measures the extent of similarity of the velocities of two nodes that are not too far away.

[7] Spatial values will small if mobile nodes independently of one another and vice-versa.

[8] Temporary dependence measures the similarity of the velocities of a node at 2 nearby time slots. Geographic restriction represent the degree of freedom of point(nodes) on a map.

4.2.1 Random Waypoint Mobility model

The simulation results (in Figure 17) show that, the energy consumption of mobile nodes in this scenario is very similar. Nodes moving randomly cause topology to change dynamically therefore the gateway discovery protocol are activated frequently. This gateway discovery protocol requires most of the nodes stay awake, thus all nodes consume more energy although the node may not have data to transmit or receive. When topology is stable ("pause time" value is long) the mobile nodes consumed less energy than when they are always moving.

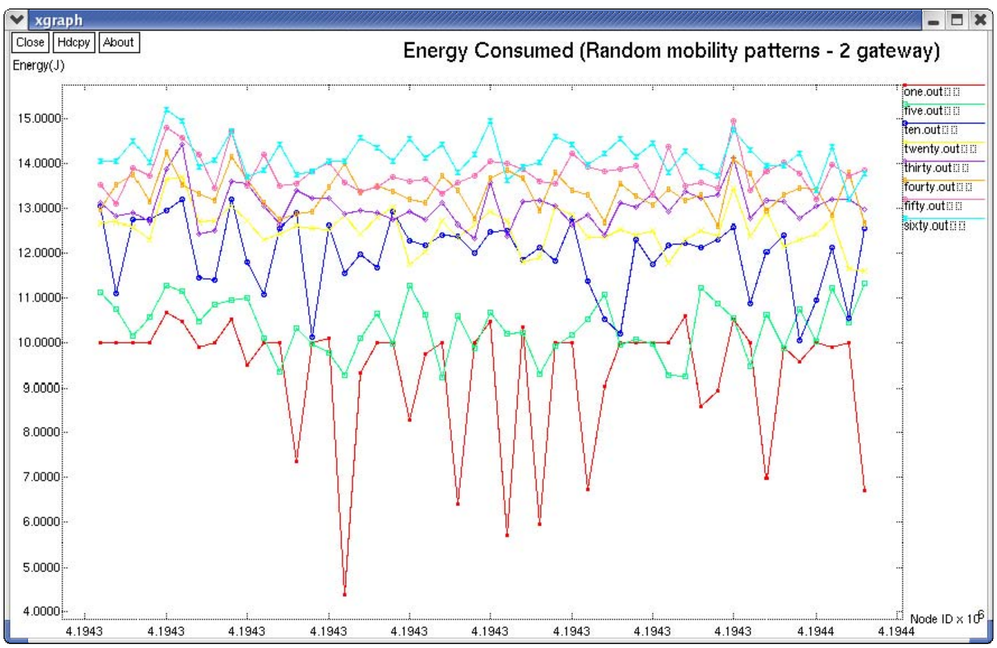

Figure 17: Energy consumed in Random Waypoint scenarios

4.2.2 Reference Point Group Mobility model

In this scenario, like the random waypoint mobility patterns, if the mobile nodes move slowly, the energy consumption is low. However, the analyzed data in Figure 18 shows that, if the topology changes frequently, several intermediate nodes will consume much more energy than others. In this scenario, these nodes are the group leader nodes.

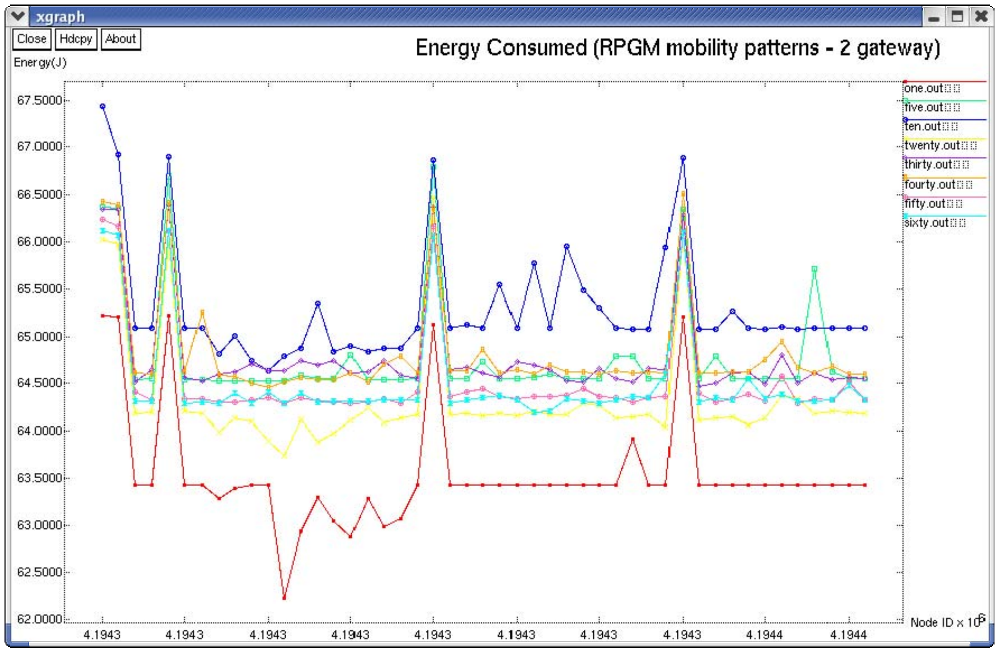

Figure 18: Energy consumed in RPGM scenarios

4.2.3 Freeway Mobility model

In this class of mobility patterns (see Figure 19), there is a big difference between the low mobility scenarios and highly mobility scenarios. When nodes move at low speed on the highway, almost all other nodes have to reduce their speed as well due to the high spatial dependence and safe separation distance that must be maintained. The default gateway to reach the internet in such scenarios rarely changes. Therefore energy is used only for forwarding packets among intermediate nodes. If the traffic source node is far from the gateway, more hops are needed to reach the gateway. However, in high mobility scenarios, when the source traffic node moves near the gateway, the number of hops still high because the information about the default gateway of these nodes is still not up-to-date. Changing the timeout and diameter parameters in the expanding-ring search algorithm of the gateway discovery protocols could help the mobile node to compensate for the problem of stale information about the default gateway. However, increasing such values will increase the overhead in the network since the gateway discover messages will be flood unnecessary, which translates into greater energy costs.

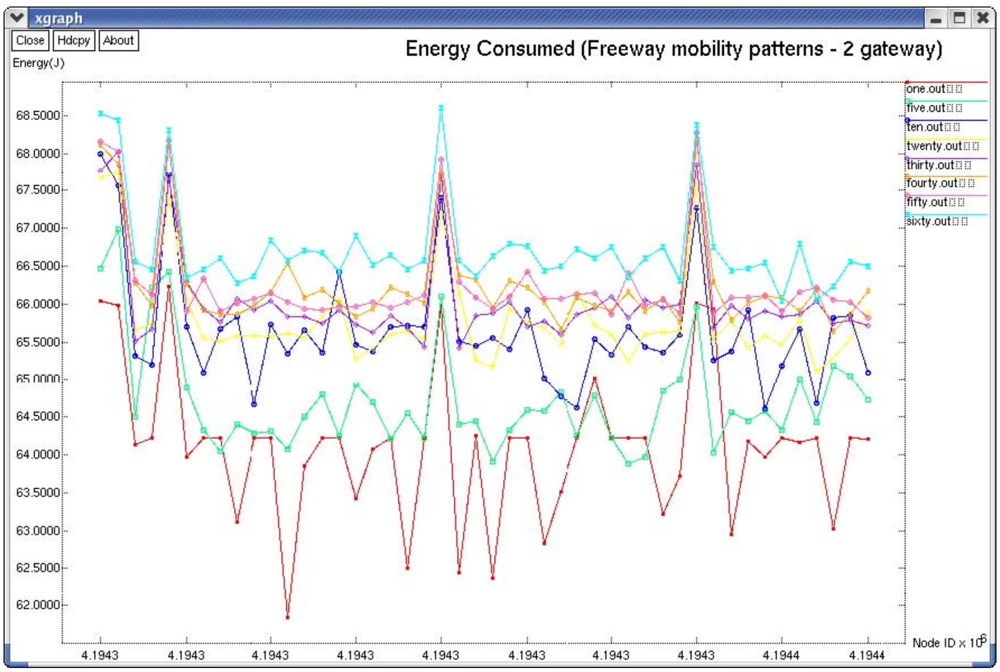

Figure 19: Energy consumed in Freeway scenarios

4.2.4 Manhattan Mobility model

The maximum and minimum energy cost of node does not show much difference (see Figure 20), the reason is that, node area distrusted randomly in the street and to reach the destination gateway, almost nodes have to travel a similar number of hops. The obstacles and the geographical limitation mean that nodes have different ways to choose to reach the gateway equally cost. Each time the gateway discover protocol is run, the node might change to a different default gateway due to a slightly movement of nodes.

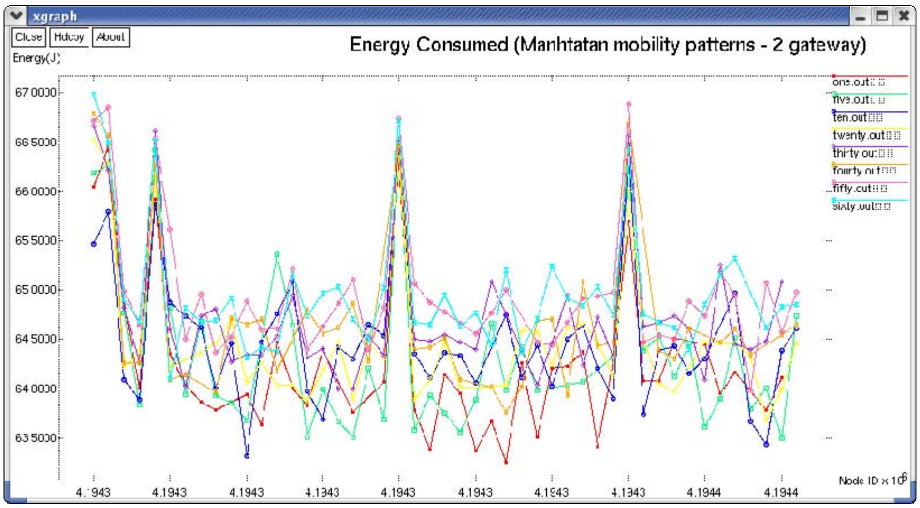

Figure 20: Energy consumed in Manhattan scenarios

The Manhattan mobility model is expected to have high spatial dependence and high temporal dependence. It too imposes geographic restrictions on node mobility. However, it differs from the Freeway model in giving a node some freedom to change its direction.

CHAPTER 5: CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

Building on the previous study [40], this thesis takes a step further by taking into consider the energy consumption in a mobile ad hoc network as a function of the number of gateways and mobility patterns.

- We have made an in-depth study concerning Internet connectivity for a mobile ad hoc network. We have evaluated a number of different scenarios and have chosen and recommend using reactive gateway discovery protocols using an IPv6 addressing scheme. This solution provides greater flexibility and scalability while providing the necessary functionality and high performance for a ad hoc network.

- To capture the interesting mobility features including spatial, temporary dependence and geographic restriction, the thesis has investigated and evaluated an ad hoc network while the internetworking with the internet under several mobility patterns (Random Waypoint, Freeway, Reference Point Group Mobility, Manhattan). The analysis and discussion of each mobility pattern with regard to energy offer some useful insight for the development and deployment of real-world scenarios.

- The simulation results show that increasing the number of gateways could significantly improve the energy saving of a mobile node in most of mobility scenarios. Although the energy efficiency is not proportional to the number of gateways, as the number of gateways change from 1-2 and from 2-3 in the tested scenario, the energy saving could change by as much as 20%.

5.2 Future work

There are many parameters that can be tuned to provide better performance and increase energy saving in these mobility patterns. Choosing the optimal parameters for a specific scenario requires in-depth study and extensive research.

When mobility is high, most of nodes stay awake longer than necessary, which results in increasing energy consumption In addition, broadcast messages used in route and gateways discovery protocol cause nodes to stay awake during the beacon interval. Hence research into mobility prediction or proactive handoffs may be needed to reduce the number of nodes that remain in an active state.

When the number of gateways reach a threshold, increasing the number of the gateway does not reduce the energy consumed of mobile nodes. In such scenarios, the coverage is entire area therefore mobile nodes could reach the gateway in one hop. However, place more gateway in such networks can increase the availability and the robustness of the network since the one or several gateway could be down or work incorrectly sometimes. Find out how many gateways should be placed for different scenarios with different mobility patterns could also be an interesting research subject.

REFERENCES

[1] S. Singh, M. Woo, and C. S. Raghavendra 1998, "*Power-aware routing in mobile ad hoc networks*." In Proceedings of the 4th Annual ACM/IEEE international Conference on Mobile Computing and Networking (Dallas, Texas, United States, October 25 - 30, 1998). W. P. Osborne and D. Moghe, Eds. MobiCom '98. ACM Press, New York, NY, 181-190.

[2] J. Li, D. Cordes, J. Zhang, "*Power-aware routing protocols in ad hoc wireless networks*" Wireless Communications, IEEE [see also IEEE Personal Communications] Volume 12, Issue 6, Dec. 2005 Page(s):69 – 81.

[3] S. Rajeev, A. Ananda, M. Chan, W. Ooi, "*Mobile, Wireless and Sensor Networks*" Wiley-IEEE Press 2006, Chapter 5, Page(s): 72-132.

[4] M. Krunz, A. Muqattash, S. Lee, "*Transmission power control in wireless ad hoc networks: challenges, solutions and open issues"* Network, IEEE Volume 18, Issue 5, Sept.-Oct. 2004 Page(s):8 – 14.

[5] S. Park, R. Sivakumar, "*Quantitative analysis of transmission power control in wireless ad-hoc networks*" Parallel Processing Workshops, 2002. Proceedings. International Conference on 18-21 Aug. 2002 Page(s):56 – 63.

[6] A. Muqattash; M. Krunz, "*POWMAC: a single-channel power-control protocol for throughput enhancement in wireless ad hoc networks*" Selected Areas in Communications, IEEE Journal on Volume 23, Issue 5, May 2005 Page(s):1067 – 1084.

[7] A. Avudainayagam, Y. Fang, and L. Wenjing, "*DEAR: a device and energy aware routing protocol for mobile ad hoc networks*" MILCOM 2002. Proceedings Volume 1, 7-10 Oct. 2002 Page(s):483 - 488 vol.1.

[8] C.-K. Toh, H. Cobb, and D.A. Scott, "*Performance evaluation of battery-life-aware routing schemes for wireless ad hoc networks,* "Communications, 2001. ICC 2001. IEEE International Conference on Volume 9, 11-14 June 2001 Page(s):2824 - 2829 vol.9.

[9] R. Zheng, R. Kravets, "*On-demand power management for ad hoc networks*, "INFOCOM 2003. Twenty-Second Annual Joint Conferences of the IEEE Computer and Communications Societies. IEEE Volume 1, 30 March-3 April 2003 Page(s):481 - 491 vol.1.

[10] C. E. Jones, K. M. Sivalingam, P. Agrawal, and J. C. Chen 2001; "*A Survey of Energy Efficient Network Protocols for Wireless Networks.*" Wireless Network. 7, 4 (Sep. 2001), 343-358.

[11] S. Singh and, C. S. Raghavendra 1998. "*PAMAS—power aware multi-access protocol with signaling for ad hoc networks.*" SIGCOMM Computer Communication. Rev. 28, 3 (Jul. 1998), 5-26.

[12]B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris 2002. "*Span: an energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks*. " Wireless Network 8, 5 (Sep. 2002), 481-494.

[13]J. Gomez, A. T. Campbell, M. Naghshineh, and C. Bisdikian 2003. "*PARO: supporting dynamic power controlled routing in wireless ad hoc networks*. " Wireless Network 9, 5 (Sep. 2003), 443-460.

[14]P. E. Charles, M. B Elizabeth, and I. Chakeres. "*Ad Hoc on Demand Distance Vector (AODV) Routing*." IETF Internet draft, Oct 2003. URL: http://moment.cs.ucsb.edu/pub/draft-perkins-manet-aodvbis-00.txt (Last accessed: 05th January 2007).

[15]B. J. David, A. M. David, and Y. Hu, "*The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks (DSR),* ",INTERNET-DRAFT 19 July 2004. (Last accessed: 05th January 2007).

[16]A. Muqattash, M. Krunz, "*Power controlled dual channel (PCDC) medium access protocol for wireless ad hoc networks*, "INFOCOM 2003. Twenty-Second Annual Joint Conferences of the IEEE Computer and Communications Societies. IEEE Volume 1, 30 March-3 April 2003 Page(s):470 - 480 vol.1.

[17]R. Ramanathan and R. Rosales-Hain. "*Topology control of multi-hop wireless networks using transmit power adjustment*." In Proceedings of the IEEE INFOCOM Conference, volume 2, pages 404–413, 2000.

[18]R. Wattenhofer, L. Li, P. Bahl, and Y.-M. Wang. "*Distributed topology control for power efficient operation in multi-hop wireless ad hoc networks*. "In Proceedings of the IEEE INFOCOM Conference, volume 3, pages 1388–1397, 2001.

[19]T. A. ElBatt, S. V. Krishnamurthy, D. Connors, and S. Dao. "*Power management for throughput enhancement in wireless ad-hoc networks*." In Proceedings of the IEEE ICC Conference, volume 3, pages 1506 - 1513, 2000.

[20]S. Khurana, A. Kahol, and A. P. Jayasumana. "*Effect of hidden terminals on the performance of IEEE 802.11 MAC protocol*." In Proceedings of the IEEE LCN Conference, pages 12–20, 1998.

[21]F. A. Tobagi and L. Kleinrock. "*Packet switching in radio channels: Part ii–the hidden terminal problem in carrier sense multiple-access and the busy-tone solution*." IEEE Transactions on Communications, 23(12):1417–1433, Dec 1975.

[22]*ISO/IEC 8802-11 IEEE Standard 802.11g* (2003) – Sections 9.2.5.6, 9.6, & 9.10. URL: http://standards.ieee.org/getieee802/download/802.11g-2003.pdf (last accessed: 6th January 2007).

[23]L. M. Feeney, 2001. "*An Energy Consumption Model for Performance Analysis of Routing Protocols for Mobile Ad Hoc Networks*." Mobile Network and Application 6, 3 (Jun. 2001), Pages 239-249.

[24]P. Ratanchandani, R. Kravets, "*A hybrid approach to Internet connectivity for mobile ad hoc networks*," Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE Volume 3, 16-20 March 2003 Page(s):1522 - 1527 vol.3

[25]T. Shojiro, S. Kaoru, and Y. Yasuhiko. "*Quick Data-Retrieving for U-APSD in IEEE802.11e WLAN Networks*," IEICE Trans A: Fundamentals, July 2006; E89-A: 1919 - 1929.
[26]S. Ray, J.B. Carruthers, and D. Starobinski, "*RTS/CTS-induced congestion in ad hoc wireless LANs*," *Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE*, vol.3, no.pp. 1516- 1521 vol.3, 16-20 March 2003.
[27]J. Shin; H. Lee; J. Na; Park, A.; Kim, S., "*Gateway discovery and routing in ad hoc networks with NAT-based Internet connectivity*," Vehicular Technology Conference, 2004. VTC2004-Fall. 2004 IEEE 60th, vol.4, Pages. 2883-2886 Vol. 4, 26-29 Sept. 2004.
[28]X. Bin; A. Kumar, "*Integrated connectivity framework for Internet and ad hoc networks*," Mobile Ad-hoc and Sensor Systems, 2004 IEEE International Conference vol. 4, pages. 540- 542, 25-27 Oct. 2004.
[29]H. Ammari, H. El-Rewini, "*Performance evaluation of hybrid environments with mobile gateways*," Computers and Communications, 2004. Proceedings. ISCC 2004. Ninth International Symposium on, vol.1, pages. 152-157 Vol.1, 28 June-1 July 2004.
[30]H. Ali, "*Internet Access and QoS in Ad Hoc Networks, Licentiate thesis*," Department of Communication Systems, Lund University, April 2006.
[31]*"IEEE P802.11e/D13.0, Part11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications: Amendment 7: Medium Access Control (MAC) Quality of Service (QoS) Enhancements*", September 2004.
[32]*"ANSI/IEEE Standard 802.11, Part11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*", 1999.
[33]*USC mobility generator*. URL: http://nile.usc.edu/important/software.htm (Last accessed: 09th January 2007)
[34]*USC ns-2 network simulator*. (version 2.30 all-in-one package) URL: http://nsnam.isi.edu/nsnam/index.php/Main_Page (Last accessed: 09th January 2007).
[35]J. McDougall; S. Miller, "*Sensitivity of wireless network simulations to a two-state Markov model channel approximation*," Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE, vol.2, page. 697- 701. Vol.2, 1-5 Dec. 2003.
[36]R. Wakikawa, J. Malinen, C. Perkins, A. Nilsson and A. J. Tuominen. "*Global Connectivity for IPv6 Mobile Ad Hoc Networks*", IETF Internet Draft, February 2003. Work in progress.
[37]The CMU Monarch Project, Wireless and mobility enhancements to ns-2, http://www.monarch.cs.cmu.edu (Last accessed: 20 January 2007)
[38]P. Johansson, T.Larsson, N. Hedman, B. Mielczarek, and M. Degermark, "*Scenario-based performance analysis of routing protocols for mobile ad-hoc networks*," in International Conference on Mobile Computing and Networking (MobiCom'99), 1999, page. 195-206.

[39]X. Hong, M. Gerla, G. Pei, and C.-C Chiang, "*A groups mobility model for ad hoc wireless networks*," In Proceedings of the 2nd ACM international Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems (Seattle, Washington, United States, August 20 - 20, 1999). MSWiM '99. ACM Press, New York, NY, 53-60.

[40]Nguyen Manh Ha and Nguyen Dinh Viet (2005) "*Performance evaluation of routing protocols in MANET*" in Fundamental & Applied IT Research Symposium, HCM city, Vietnam 09/2005. Sponsored by College of Applied Science and Technology.

[41]P. Hurni, T. Braun, and L. M. Feeney. "*Simulation and evaluation of unsynchronized power saving mechanisms in wireless ad hoc networks*. " In 4th Internationl Conference on Wired/Wireless Internet Communications WWIC 2006, Bern, Switzerland, May 2006.

[42]Devin Akin "*Protection Ripple in 802.11 WLANs*." 2004 White page Planet3 Wireless, Inc. Pages 3-7.

[43]J.-H. Chang and L. Tassiulas, "*Energy conserving routing in wireless. ad-hoc networks*," in Proc. IEEE INFOCOM, pp. 22–31, March 2000

APPENDIX A

A.1. Setup Internet Connectivity implementation for MANET in ns-2

The core part of the module based [30]. We have added and edited the parameters used in the gateway discovery process to reflect the particular scenarios. The version using for this thesis can be downloaded from: http://web.it.kth.se/~nguyenh/master_thesis/code/manet_I.tar.gz.

```
1. Download manet_I.tar.gz to the “aodv” directory
2. Backup your original aodv directory. The files extraction
process will overwrite all files in aodv directory
3. untar –zxvf manet_I.tar.gz
4. Go to /tcl/lib directory (cd ../tcl/lib)
5. In ns-default.tcl, add this (preferably under "Agents"):
        Agent/AODV set gw_discovery 2
6. Recompile ns2
        #cd ../..
        #make clean
        #make
```

A.2. Mobility scenario generation

A.2.1 Random Waypoint model

This mobility model is included with ns2 release as application named setdest. To use this utility go to the ns2 directory and compile it:

```
#cd indep-ultil/cmu-mobi-gen/setdest
#make all
```

A.2.2 Reference Point Group Mobility (RPGM), Freeway Mobility (FW), Manhattan Mobility (MH) model are delivered in separated packets. These tool need to be edit before working properly. We have created the makefile for all packages. The ready to run version can be downloaded from the follow URL: http://web.it.kth.se/~nguyen/master_thesis/code/mobigen.tar.gz. Extract the compressed file and use make to compile the application.

```
#tar –zvxf mobigen.tar.gz
#cd mobigen
#make all
```

A.3. Simulation code

It is required that you install above modules (Internet connectivity for MANET and Mobility Generator) before running the simulation scripts. The simulation code used for simulation in the thesis can be downloaded from: http://web.it.kth.se/~nguyenh/master_thesis/code/simulation.tar.gz. Extract files and run as normal ns2 script.

A.4. Post processing

Result of the simulation is a collection of file called traced file. In order to get the result and meaningful presentation, post processing is needed. Post processing script will analyze trace file and extract information based on the desired metric. The source code of those scripts can be downloaded from URL: http://web.it.kth.se/~nguyenh/master_thesis/code/post_processing.tar.gz.

COS/CCS 2007-06

www.kth.se