A New Routing Strategy for Mobile Ad Hoc Communication

Anders Lundström
Magnus Westbergh

Submitted for the Degree of
Master of Science in Electrical Engineering

Department of Technology
University of Kalmar
S-391 82, Kalmar, SWEDEN

June 2004

Abstract

With a human necessity to be able to communicate with people in their proximity from anywhere to anyone ad hoc networks can be used to instantly connect to local or remote networks such as the Internet without the need of pre-existing infrastructure. Without the need of centralized administration or pre-existing infrastructure, users of the network together establishes the infrastructure.

Wireless communication has one great disadvantage and that is the limited range of radio transmissions. An elegant solution is to use ad hoc networking in which data can be forwarded by intermediate systems through the network in able to reach the final destination. Such a network can be used in many different applications, for instance spontaneous and temporary surveillance systems in hazardous environments.

An ad hoc network consists of arbitrary deployed communicational devices, such as cellular phones, Personal Digital Assistants (PDAs) etc.

This thesis was aiming to develop a new routing strategy for mobile ad hoc communication with key features such as utilization of as little computational resources as possible. Thereby utilize less consumed power due to the fact that most wireless devices use batteries and therefore have a limited lifetime. Requirements were made that routes should be established as fast as possible to provide almost instant communication. Stability issues such as link connectivity between intermediate systems have also been an important issue during the development.

Key words: Ad Hoc, Mobile Ad Hoc Network., MANET, Self configuring, ARP, Address Resolution Protocol, Networking, Wireless LAN, IEEE802.11b

Email: anders@comcon.se
magnus@comcon.se

WWW: http://www.comcon.se

Acknowledgments

We would like to give an acknowledgement to all the people who made this master thesis possible. It has been a great pleasure to work with the people at Ericsson Microwave Systems AB.

Thank you all!

Contents

Chapter 1

Introduction

1.1 Background

During the last decades, wireless communication strategies have become more crucial for mankind. Devices such as cellular phones have become property of each and everyone, with urban coverage nearly anywhere in the world, allowing people to communicate with each other. There is a technological revolution present, where wireless communication and wireless communicating devices are the key elements. Technologies such as Bluetooth™ and WLAN have become available, facilitates easier living and creating the wireless environment that many people talk about.

Ad hoc communication is a technology that provides the possibility for devices to communicate with each other through each other. Intermediate devices between the source and the destination will act as relay stations. The ad hoc technology provides benefits such as reduction of transmission output and thereto decreased battery consumption. The characteristic of an ad hoc network is a network that makes pre-existing infrastructure obsolete and with technology that provides dynamic topology. Further, the ability of a self healing structure makes the communication less vulnerable for failing links. That is, communicating devices may be removed or added to the network; still the information will make its way through the network to its final destination.

The existence of ad hoc technology will not solve all communicating problems; however, small to medium networks could use the technology with advantage. Existing products on the market using limited variants of ad hoc are amongst others, Bluetooth™ and WLAN.

The field of ad hoc networking has during the last decades become a subject to a lot of interest. The ad hoc technology is amongst other technologies supposed to be a vital part of future military application, Network Centric Warfare (NCW).

The possibilities of usage in civilian applications seem to be infinite. For instance, wireless intrusion alarms, chemical sensing in industrial environments, traffic surveillance, catastrophic aid networks, cellular phones and cellular base transceiver stations.

For military purposes, applications such as Unattended Ground Sensor (UGS) networks for reconnaissance and surveillance can be based on ad hoc technology to provide quick deployment in hostile environments.

Generally, great benefits can be made in any network that demands rapid deployment and used during a limited time period. However, this does not exclude the usage of ad hoc in static network environment.

The ad hoc technology is still in its inception, however, much research is being done within the field and it will be more common within the next decades.

1.2 Problem Description

The main purpose of this master thesis was to develop a new strategy for ad hoc routing, utilizing less overhead network traffic than already existing protocols along with utilization of less computational resources. Further, the routing protocol should take link quality into consideration to provide reliable communication links. Efficiency and simplicity has been two main criteria of this work.

1.3 Project Organization

The following persons have been involved in this master thesis:

Thesis authors

Anders Lundström
Magnus Westbergh

Supervisor at Ericsson Microwave Systems AB

Leif Axelsson, Ph.D.

Examiner at the University of Kalmar, Department of Technology

Professor Wlodek Kulesza

Hardware and Linux Support at Ericsson Microwave Systems AB

Benny Sjöstrand
Henrik Rundqvist

1.4 Disposition

The work described in this master thesis is structured as follows.

Firstly, a background of the subject and the theories in general of ad hoc networking is presented. Section 2.2 and 2.3 present different routing strategies and existing protocols.

In chapter 3 the Open Systems Interconnection (OSI) Network Reference Model is briefly described to ease further understanding about packet based communication such as TCP/IP.

The Address Resolution Protocol (ARP) presented in chapter 4, describes the protocol and its purpose in networks such as the Internet.

Chapter 5 and further describes the theories of Multi-hop Enabled ARP (MEARP). Issues dealt within this section are the concept of the routing algorithm, route management and link quality aspects. Section 5.5 describes how MEARP has been successfully implemented and some issues for further development.

Chapter 2

The Concept of Ad Hoc Routing

2.1 Introduction to Ad Hoc Routing

In Latin, ad hoc literally means “for this”, and is often applied in terms of networks where devices may be added and integrated using wireless technology. As a definition, an ad hoc network is one that comes together as needed, not necessarily with any assistance from the existing Internet infrastructure, simply a collection of wireless mobile hosts forming a temporary network without centralized administration [1].

In a regular wired network of today, such as Internet, transmitted data packets are forwarded from source to destination by routers defined with fixed rules. Such a network consists of mostly predetermined infrastructures. Each node has been manually configured to its present state. In this type of network all communication depends on a centralized administrative infrastructure, unlike an ad hoc network.

Wireless data communication operates in two different modes. The infrastructure mode, where all the communication goes through an access point. An access point connects users to other users within the network and also can serve as the point of interconnection between the WLAN and a fixed wired network, i.e. the Internet. The second mode is the peer-to-peer mode, where the nodes can communicate directly if they are within range of each other.

A Mobile Ad hoc Network, referred to as MANET is an autonomous multi-hop system of mobile hosts connected by wireless radio frequency (RF) links. If two hosts are not within radio range, all data packets must pass through intermediate hosts which doubles as routers, and performs the same job as routers within the Internet infrastructure [2] [3].

A parallel can be made between MANET communication and the way humans relay information to each other. Imagine standing in a crowded square wishing to speak to your friend far away in the crowd. Instead of shouting out your message, you notice the people surrounding you. You may now ask another person to relay your message. The message will traverse its way through the crowd and finally reach your friend. In the same moment you have transformed the crowd to a human variant of an ad hoc network.

Ad hoc routing is what underlies the establishment of the paths by which the MANET nodes can communicate with each other. The routing maintains the routes and makes it transparent to the user.

2.2 Characteristics

A MANET consists of mobile platforms (e.g., a router with a wireless communication device), which are free to move about arbitrarily. These nodes may be located in or on airplanes, ships, cars, or even on people or very small devices, and there may be multiple hosts per router. A MANET is an autonomous system of mobile nodes. This system may operate in isolation, or may have gateways to and interface with a fixed network as well as other MANETs [1] [2].

The difference between a wired static network and a MANET is great in many ways. There are some key challenges associated with an ad hoc environment, and some specific characteristics. Some of the most salient characteristics will be recognized in this section.

Dynamic topology: Nodes within the ad hoc network may move arbitrarily and randomly. The routing between the different nodes often consists of multi-hop, therefore routes may change hastily and irregular. Due to the nodes movement, the routing information will need to be updated more often than its wired counterpart. This results in more routing overhead information, which in turn increases the use of the radio medium resources. This puts great emphasize in choosing routing protocol, when all the different protocols uses different strategies to establish and maintain routes.

Multi-hop: The communication between the nodes in an ad hoc network consists of wireless transmissions, where all the nodes most likely not have a peer-to-peer contact with each other. The connection is not guaranteed. The data packets have to be forwarded between multiple intermediate nodes to reach the destination. In a MANET the topology may change constantly, due to possibility of node movement.

Bandwidth-constrained, variable capacity links: As a contrast to hardwired networks, communication over wireless medium will continue to have less bandwidth capacity. In addition signal interference, noise and fading, will most likely be higher, resulting in even less useful bandwidth. Congestion is very common, i.e. aggregate applications demand, will likely approach or exceed the network capacity, as the mobile networks often is simply an extension of a fixed infrastructure.

Energy-constrained operation: Most of, or all nodes in a MANET may relay on batteries or other consumable means for their energy. The most important system design for these nodes is criteria for optimization and energy conservation.

Limited physical security: Ad hoc networks are generally more vulnerable to physical security threats than fixed- cable nets. There is a great threat of possible strikes against the network, such as eavesdropping, spoofing, and denial-of-service attacks, which should be carefully considered. To reduce security threats existing link security techniques can be applied within wireless networks. As a benefit, the decentralized nature of network control in MANETs provides additional robustness against the single point of failure of more centralized approaches.

These characteristics create a set of underlying guidelines and performance concerns for protocol design, which extends beyond those guiding the design of routing within the higher-speed, semi-static topology of the fixed Internet.

2.3 Conceivable Usage Areas

During the last years mobile computing has grown in popularity. At the same time, the markets for wireless telephones and communication devices are experiencing rapid increase. Much of the interest has to do with keeping in touch with the Internet. We expect to have the network at our disposal at all times. We might wish to download a map on the run so that we can see what is available in the local area. We may want to have driving suggestions sent to us, based on information from the Global Positioning System (GPS) in our car.

As wireless network nodes increases and as applications using the Internet becomes familiar to a wider class of customers, the customers will expect to use network applications even in situations where Internet itself is not available. For instance, people using laptop computers at a conference in a hotel might wish to communicate in a variety of ways, without the mediation of routing across the global Internet. Today such obvious communications requirements cannot be easily met using Internet protocols. Ad hoc provides solutions to meet such requirements and in this section we will look at some of the potential applications and usage areas for ad hoc networks.

Emergency services: Today, extreme terrorists terrify the globe and the nature has become wild and dangerous. Every year earthquakes, massive storms, and other natural disasters occur. What happens when our existing infrastructure is damaged or out of service for some reason? As the Internet grows in importance, the loss of network during such a catastrophe will have a noticeable significance. A communication network might break down, and what happens when no phone-calls can be made?

Ad hoc networks can help to overcome network impairment during disaster emergencies. Mobile units can carry networking equipment in support of routine operations for the times when the Internet is available and the infrastructure has not been impaired. With ad hoc techniques, emergency mobile units can function as an emergency network and for instance, help police, and firefighters to stay in touch and provide information more rapidly.

Conferencing: When mobile computer users gather outside their normal office environment, the business network infrastructure is often missing. But the need for shared computing might be even more important here than in the everyday office environment. This can be solved by the establishment of an ad hoc network for collaborative computing.

Cellular network: If a user of a cellular phone is out of reach from a base station normally he/she will not be able to use the device properly. This problem can be solved if there were other intermediate users between the end user and the base station. With ad hoc technology implemented, the out of reach user could get connection through multi-hop. This would solve tremendously many situations of today when you cannot use your cellular phone due to bad reception. The technical implementation can be solved. But how will the intermediate node users be compensated for relaying data and the use of their own resources, such as battery?

Military sensor networks: The sensor network is a surveillance system that may be a complement to land mines e.g. the enemy is spotted without him knowing about it. From a network of sensors, target detection, tracking, localization, and recognition are vital information that can be determined. There is a need to deploy several sensor nodes in the vicinity to ensure continuous monitoring of detection. A network of nodes that uses multiple sensor technologies can accurately locate and identify targets in the area. By placing the sensors in an ad hoc network, many advantages will be achieved such as the possibility to a highly redundant network and possibility to add more sensors to the network if needed [4].

Military soldier networks: To ensure safety and accuracy on battlefields and in hostile environments, communication is crucial. To be able to act in best way you need a clear image of what occurs in the nearby region. The one with information superiority will have the greatest prerequisite of succeeding. By introducing ad hoc communication between soldiers and vehicles, information can be shared and situation pictures can be created from the information received from the different locations. Commanders can easily brief all available personal, by transmitting orders in the network. The ad hoc network relays the information and provides the communication.

Embedded computing applications: The world is full of machines that move and future intelligent mobile machines will be able to process more information about the environment in which they operate. Present intelligent internetworking devices that detect their environment interact with each other, and respond to changing environmental conditions will create the future.

There are infinite of different possible applications and usage areas, both civilian and military, where ad hoc technology can be used with great benefits, it is only our imagination that stops us from finding new ones.

2.4 Routing Management

The concept of routing can be described as the process of path finding. A router is a device or in some cases, software in computers, that determines the next network point to which a data packet should be forwarded toward its destination.

As pointed out in chapter 2.1.1 an ad hoc network has some very specific characteristics compared to its hard-wired counterpart. When developing ad hoc routing algorithms, the protocols are designed to be optimal in different situation, i.e. in a mobile network it might be good to update the routes more often.

This chapter focuses on some types of classification and different features for ad hoc routing.

Proactive routing algorithms: Proactive routing, also known as table-driven routing acts similar to traditional fixed network routing. Proactive routing attempts to maintain routes to all destinations at all times, regardless of whether they are needed. To support this, the routing protocol propagates information updates about a network's topology or connectivity throughout the network. Information updates can be topology-driven, which are generated when connectivity in the network is detected; periodic, which generates connectivity information at fixed intervals; or both. Proactive routing is fast when you need a path. The protocol stores all routes in its routing table. This method is resource demanding in terms of batteries, CPU usage, power and bandwidth. Proactive routing is mostly used in mobile networks, where the routes are changed constantly, and connectivity is needed. Examples of proactive routing protocols are the Topology Dissemination Based on Reverse-Path Forwarding (TBRPF) protocol [5] and the Optimized Link State Routing (OLSR) protocol [6].

Reactive routing algorithms: Reactive or on-demand routing protocols determine routes only when there is data to send. If a route is unknown, the source node initiates a search to find one, which tends to cause a traffic flow as the query is propagated through the network. Nodes that receive the query and have a route to the requested destination respond to the query. In general, reactive protocols are primarily interested in finding any route to a destination, not necessarily the optimal route. Data sent in networks using reactive protocols do tend to suffer a delay during the search for a route. These protocols are power saving and quiet. A reactive protocol is mostly suited for slow moving and static environments, i.e. a surveillance network. Ad Hoc On-Demand Distance Vector (AODV) [7] and Dynamic Source Routing (DSR) [8] are examples of reactive routing protocols.

Hybrid routing algorithms: A hybrid routing protocol makes use of both proactive and reactive routing techniques. In limited regions around a node it may act proactive, and else reactive. If the network is moving, the protocol might sense the movement and switch over from reactive to proactive routing. By using a mixture of the two main theories, the cost of resources can be reduced. This can give more scalability to the network, but also more complexity to the algorithm. The Zone Routing Protocol (ZRP) [9] is one example of a hybrid ad hoc routing protocol.

Symmetric / asymmetric routing: This characteristic is independent of the usage of proactive or reactive routing. Symmetric routing means bi-directional link communication. A route between source and destination is the same in both directions. The data exchange between two neighbor nodes works equivalent in both directions.

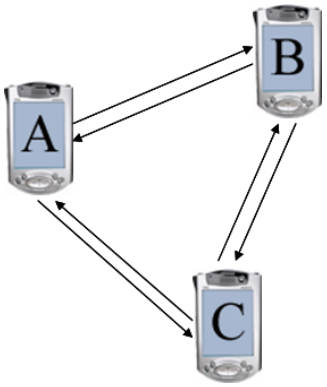

Figure 1. Bi-directional link

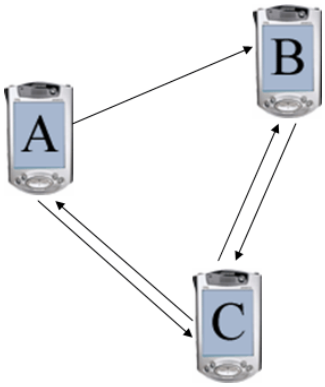

Figure 2. Uni-directional link

Routing protocols with support of unidirectional links offers the possibility of asymmetric routing. This can be useful when connectivity between two nodes differs, e.g. due to differing antenna, sources of interface or transmission power.

In real applications it is highly likely that there are some nodes in the network whose radio transceivers have higher power capacity than the other nodes. These could for instance be devices mounted on people or vehicles. Those could be used for transmissions over larger areas of the network and to more distant nodes. If the receiving nodes do not have the same transmit power, they will not be able to send packets back directly to the source. With asymmetric routing this would not be a problem.

2.5 Existing Strategies for Ad Hoc Routing

The abbreviation MANET mentioned earlier is also associated with the Internet Engineering Task Force (IETF) MANET Working Group. "The purpose of this working group is to standardize IP routing protocol functionality suitable for wireless routing application within both static and dynamic topologies. The fundamental design issues are that the wireless link interfaces have some unique routing interface characteristics and that node topologies within a wireless routing region may experience increased dynamics, due to motion or other factors" [3]. Currently the workgroup is working with standardization of four ad hoc routing protocols; DSR, AODV, OLSR and TBRPF.

Ad Hoc On-Demand Distance Vector (AODV): AODV is a reactive routing protocol, intended for use by mobile nodes in an ad hoc network. The primary goal with AODV was to reduce the routing overhead in the network as much as possible. If a node wants to know a route to a given destination it generates a Route Request (RREQ). The RREQ is forwarded by intermediate nodes which add a reverse route for itself from the destination. When the destination is found a Route Reply (RREP) is sent back to the originator. Whenever a route is available between source and destination, AODV does not add any overhead to the packets carrying the data. AODV use Hello-messages to keep track of its neighbors at all time. When routes are not used, they are expired and therefore discarded, which reduces the effect of old routes as well as the need for route maintenance for unused routes. When a route fails, the source is notified with a Route Error (RERR) and a new RREQ can be initiated if necessary. AODV use destination sequence numbers to ensure a loop freedom at all times [7].

Dynamic Source Routing (DSR): DSR is a simple and efficient routing protocol designed specifically for use in multi-hop wireless ad hoc networks of mobile nodes. DSR allows the network to be completely self-organizing and self-configuring, without the need for any existing network infrastructure or administration. The protocol is composed of the two mechanisms of Route Discovery and Route Maintenance, which work together to allow nodes to discover and maintain source routes to arbitrary destinations in the ad hoc network. The use of source routing allows packet routing to be trivially loop-free, avoids the need for up-to-date routing information in the intermediate nodes through which packets are forwarded, and allows nodes forwarding or overhearing packets to cache the routing information in them for their own future use. All aspects of the protocol operate entirely on-demand, allowing the routing packet overhead of DSR to scale automatically to only that needed to react to changes in the routes currently in use [8].

Optimized Link State Routing (OLSR): OLSR is an optimization of the pure link state algorithm tailored to the requirements of a mobile wireless LAN. OLSR is using two different message types, the Hello-, and the Topology Control-message (TP). The key concept used in the protocol is that of Multipoint Relays (MPRs). MPRs are selected nodes which forward broadcast messages during the flooding process. This technique substantially reduces the message overhead as compared to pure flooding mechanism where every node retransmits each message when it receives the first copy of the packet. In OLSR, information flooded in the network through these MPRs is also about the MPRs. Thus a second optimization is achieved by minimizing the contents of the control messages flooded in the network. Hence, as contrary to the classic link state algorithm, only a small subset of links with the neighbor nodes is declared instead of all the links. This information is then used by the OLSR protocol for route calculation. As a consequence hereof, the routes contain only the MPRs as intermediate nodes from a source to a destination. OLSR provides optimal routes (in terms of number of hops). The protocol is particularly suitable for large and not too dense networks as the technique of MPRs works well in this context [6].

Topology Dissemination Based on Reverse-Path Forwarding (TBRPF): TBRPF is a proactive, link-state routing protocol designed for use in mobile ad-hoc networks. TBRPF has two modes: full topology (FT) and partial topology (PT). TBRPF-FT uses the concept of reverse-path forwarding to reliably and efficiently broadcast each topology update in the reverse direction along the dynamically changing broadcast tree formed by the min-hop paths from all nodes to the source of the update. TBRPF-PT achieves a further reduction in control traffic, especially in large, dense networks, by providing each node with the state of only a relatively small subset of the network links, sufficient to compute minimum-hop paths to all other nodes. In both the FT and PT modes, a node forwards an update only if the node is not a leaf of the broadcast tree rooted at the source of the update. In addition, in the PT mode, a node forwards an update only if it results in a change to the node's source tree. As a result, each node reports only changes to a relatively small portion of its source tree [5].

Chapter 3

OSI Network Reference Model

3.1 OSI Reference Model

The Open System Interconnection (OSI) reference model is a conceptual model describing how information from a software application in one computer moves through a network medium to a software application in another computer [10].

In the 1980's ISO developed the seven layered reference model, which aim to guide product implementers so their products will consistently work with other products. Each layer describes different stages the information has to pass on its way from one computer to another. The layers prepare the data packets for under- or overlying layers in order to transmit or receive information.

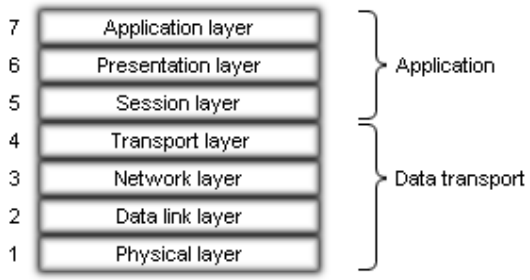

Figure 3. The OSI reference model.

The OSI reference model contains seven independent layers, which can be divided into two categories - upper and lower layers.

The upper layers of the model deal with application issues, whereas the lower layers handle data transport issues. The application layer is closest to the end user and the physical layer is responsible for actually placing the information on the medium such as network cabling.

3.2 Network Layers

3.2.1 Physical Layer

The lowest layer of the model conveys a stream of bits through the network at the electrical and mechanical level. It defines the electrical, mechanical, procedural and functional specifications for activating, maintaining and deactivating the physical link between communicating network systems. I.e. the hardware means of sending and receiving data on a carrier.

3.2.2 Data Link Layer

The data link layer controls the transmission of blocks of data over a physical network link and thereby provides a reliable transit of data. It has been subdivided by the Institute of Electrical and Electronics Engineers (IEEE) into two sub layers; Logical Link Control (LLC) and Media Access Control (MAC).

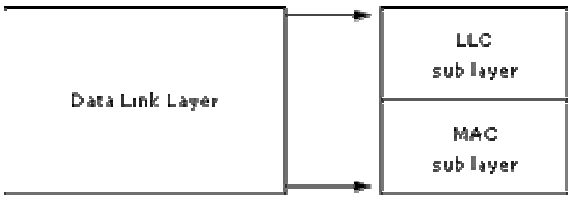

Figure 4. Sub layers of the data link layer.

3.2.2.1 Logical Link Control

The Logical Link Control (LLC) manages communication between devices over a single link of a network. It provides the ability to detect and correct errors that may occur in the physical layer. The LLC is defined in the IEEE 802.2 specification and supports both connectionless and connection-oriented services used by higher-layer protocols. The specification defines a number of fields in data link layer frames, which enables multiple higher-layer protocols to share a single physical data link, see chapter 3.2.4.

3.2.2.2 Media Access Control

The Media Access Control (MAC) layer is concerned with sharing the physical connection to the network among several computers. The IEEE MAC specification defines MAC addresses, which enables multiple devices to uniquely identify one another at the data link layer.

3.2.3 Network Layer

The network layer defines network addressing, which differs from the addressing at the data link layer. The Address Resolution Protocol (ARP), see chapter 4, is used to map the different addresses. The layer also defines the logical network layout; therefore routers can use this layer to determine how to forward data packets.

Among existing protocols that generally map to the OSI network layer are the Internet Protocol (IP) and Internetwork Packet Exchange (IPX).

3.2.4 Transport Layer

The transport layer determines through error checking, error recovering and flow control whether all data packets have arrived. It thereby ensures complete data transfer.

There are two types of transport layer protocols; connection-oriented such as the Transmission Control Protocol (TCP) and connectionless protocols such as the User Datagram Protocol (UDP). Connection-oriented protocols provide a reliable end-to-end connection between two communicating computers with packet acknowledgment and resending as two of its key features. Connectionless protocols however, are used for real-time data such as audio and video. Since the data is being processed in real-time, neither resending nor error checking is usually performed.

3.2.5 Session Layer

This layer provides the control for managing communications; for example establishment, management and termination of connections.

3.2.6 Presentation Layer

The presentation layer provides the functionality to ensure that transmitted information from the application layer of one system would be readable by the application layer of another system.

Data encryption and data compression are some of the services provided by the presentation layer to protect unauthorized access and reduction of the number of packets required for transport.

3.2.7 Application Layer

This layer is closest to the user, which means that both the application layer and the user interact directly with the software application. Services such as identification of communication partners are also done by the application layer, e.g. making sure that the other party is identified and can be reached.

Chapter 4

ARP – Address Resolution Protocol

4.1 The Internet Standard

The Address Resolution Protocol (ARP) is a method of converting link layer addresses, e.g. MAC addresses, from its network layer addresses, e.g. IP addresses. The protocol has been standardized by the Internet Engineering Task Force (IETF) in Request for Comments (RFC) 826 [11].

Today, ARP is almost exclusively used to resolve IP addresses to MAC addresses; however it was not originally intended or designed to be an IP-only protocol.

The MAC address, or physical address, is most typically a 48-bit long unique hardware address programmed into a network media. The current IP standard used on the Internet, version 4, uses 32-bit addressing.

To prevent network flooding, the ARP uses a cache to store mapped link layer addresses and network layer addresses. This is used with the assumption that MAC and IP addresses rarely change and therefore transmission of ARP messages is considered as unnecessary. The ARP cache entries have a preconfigured timeout, which allows the ARP cache to remove entries that are not in use or have been changed. The purpose of the ARP cache is to allow communicating devices to start communicating faster, without interference of ARP messages and thereby utilize less network resources [12].

An ARP message is embodied in a packet format as follows. Notice that Figure 5 is schematic with no reference to its actual size, due to variable size depending on hardware and protocol used. Since the most common usage of ARP is converting IP addresses to MAC addresses, the most typical size of an ARP message is 28 bytes.

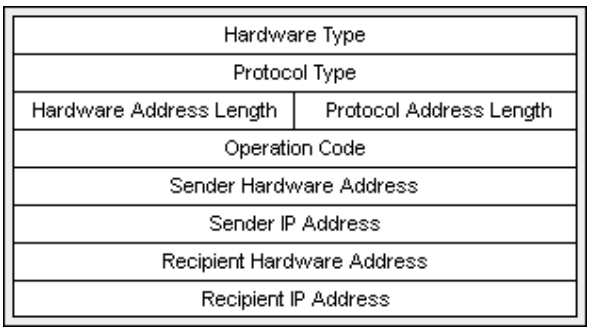

Figure 5. The Address Resolution Protocol packet format.

The most significant field in the packet format is the operation code, which allows different message types using the same packet format. Every message type uses a unique operation code.

4.1.1 ARP Request and Reply

Two of the most widely used message types are the ARP Request and the ARP Reply messages. The purpose of the ARP Request message is to request other computers MAC addresses by only knowing their network addresses, IP address that is. The ARP Reply is the recipient computers mean to answer a request.

ARP Requests are broadcasted due to the fact that the recipient hardware address is not known, therefore requested.

Figure 6. ARP resolves MAC addresses.

To illustrate how ARP resolves a MAC address the figure above is used. Node A and node B have no prior information regarding each other in their ARP cache; however node A wishes to communicate with node B.

1. Node A checks its ARP cache and realizes that an ARP Request has to be transmitted to resolve node B's hardware address.
2. Node A transmits an ARP Request.
3. Node B determines that the requested IP address matches its own and transmits an ARP Reply to node A.
4. Node A receives the ARP Reply from node B and updates its ARP cache.

Some implementations of ARP allow node B in step 3 above to update its ARP cache with a mapped hardware/software address for node A. This cause of action optimizes the usage of ARP and allows communication between node B and A to start without transmitting a request.

Once the MAC address of node B is determined, IP based traffic from node A to node B may precede.

Chapter 5

Multi-hop Enabled ARP

5.1 The Concept of Multi-hop Enabled ARP

Wireless communication between mobile users is becoming more and more popular as devices and technology is being developed. The concept of ad hoc becomes more familiar to people and an abundance of different ad hoc routing protocols are being developed. Common for most of the protocols are that they operate on level 3 in the OSI reference model, see chapter 3.2.3. When routing on level 3 in the OSI reference model the received data must be processed in several stages which requires batteries and data power.

MEARP is a self-configuring routing protocol without any need of pre-existing infrastructure. The keyword for Multi-hop Enabled ARP (MEARP) is simplicity. MEARP is a resource-less routing protocol. It operates as a reactive routing algorithm, on layer 2.5 in symbioses with the Address Resolution Protocol (ARP), between the data link layer and the network layer. By enabling the routing at a lower level process time will decrease and battery power will be saved.

In order to establish routes MEARP uses and reuses existing network traffic to minimize the overhead traffic in the network. The routes are discovered with assistance from ARP messages. By using ARP messages for route discovery MEARP becomes fully compatible to other wireless systems that do not support the routing algorithm. This enables peer-to-peer communication between a MEARP node and a common wireless computer system.

To maintain routes and keep the links reliable, the link quality and other cross-layer issues are considered. By measuring the link quality, new routes can be created in advance, before an old link is lost. This assumes that there exists an alternative route.

5.2 ARP Messages for Ad Hoc Purposes

The most common way for ad hoc routing protocols to create routes is to send a route request, which acts like ARP requests, with minor modifications. A new approach is to retransmit ARP request over the network and thereby creating an ARP initiated route request. To make this work, the retransmitting node has to change the ARP request by setting itself as the sender and remember who was asking for whom, see chapter 5.3.2.

A suitable solution for forwarding of IP packets would be to use a technique called IP forwarding, which primarily is used by routers on the Internet. IP forwarding takes care of framing of the data with a data link header containing the data link layer destination address of the next hop along a path towards its destination.

5.2.1 ARP Route Forward

The ARP Route Forward message type has been created to allow intermediate nodes to send ARP replies and to make the recipient aware of both the destination and the intermediate node.

Imagine a scenario with three nodes as shown in figure 7.

Figure 7. ARP Request in an ad hoc environment.

In this example, nodes can only communicate with adjacent nodes, which make it impossible for node A to communicate directly with node C. However, by allowing node B to retransmit ARP requests initiated by node A and transmitting an ARP Route Forward message to node A when receiving an ARP Reply from node C, a route is established between nodes A and C.

The forward message contains information about the IP address and the MAC address to the destination, that can be reached, and what node that sends the forwarding message. In this example, the forward message from node B contains both node B's and C's IP and MAC address, and it is sent to node A. From node A's point of view, messages to node C should be sent to node B to be able to communicate.

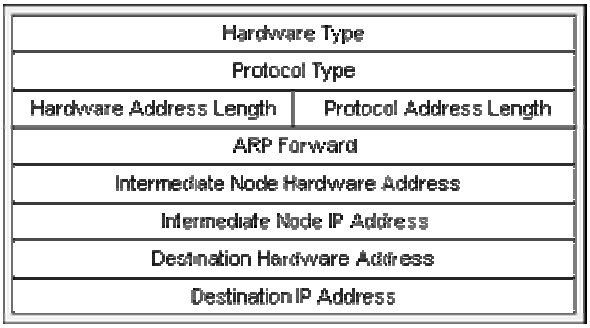

Figure 8. The ARP Route Forward message header.

Figure 8 shows how the ARP Route Forward message is structured. A new operation code has been introduced to make it possible to identify this message. A system without the ability to interpret this message will discard the message due to the ARP specification.

5.2.2 ARP Route Error

The ARP Route Error message type has been created to allow intermediate nodes to tell other nodes that a connection to a third node has been terminated, disconnected or unavailable. The route error message contains information about which node that can not be reached and which node that transmits the route error. This allows recipients to determine whether they have any alternative routes towards the destination or if they should send an ARP Route Error by themselves. New routes are discovered by the ARP Request mechanism, if needed.

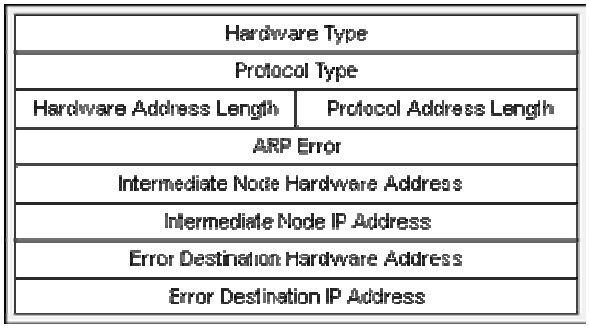

Figure 9. The ARP Route Error message header.

Figure 9 shows the structure of the ARP Route Error message. This message contains a new operation code to allow interpretation, analogous as the ARP Route Forward message.

5.3 Route Management

As described in chapter 5.1, the strategy of MEARP is to let communicating devices rebroadcast incoming ARP requests and replies. However, this feature requires that intermediate nodes stores information about sent and received ARP messages. To solve this issue, a pending list is being introduced.

The purpose of rebroadcast incoming ARP requests is to find a path to the destination through intermediate nodes. Normally, in non ad hoc networks, only adjacent nodes receive the broadcasted ARP request.

5.3.1 Route Table

One of the key components of MEARP is the internal routing table, which contains information about all known destinations and their first intermediate node, usually called gateways.

An optional feature of MEARP is to allow the routing table to contain redundant paths towards a certain destination. There are basically two advantages of using this feature:

1. Divide the traffic between the paths, usually referred to as multi-path, to utilize less network resources over the same path.
2. Allow quick reroute in a more or less static ad hoc network, when multi-path is not used, thanks to the fact that links rarely changes. However, in a more mobile ad hoc network, this feature can reduce the rerouting efficiency of MEARP.

Another optional, but highly recommended, feature is the route timeout feature. The purpose with the route timeout is to automatically remove any unused route from the routing table. In the case of a more or less static ad hoc network a longer timeout could be chosen thanks to the fact that network topology rarely changes. However, in a highly mobile network, the timeout should be set more wisely.

Figure 10 shows a six node scenario. The figure schematically shows how the nodes relate to each other geographically based on the routing information displayed in figure 11.

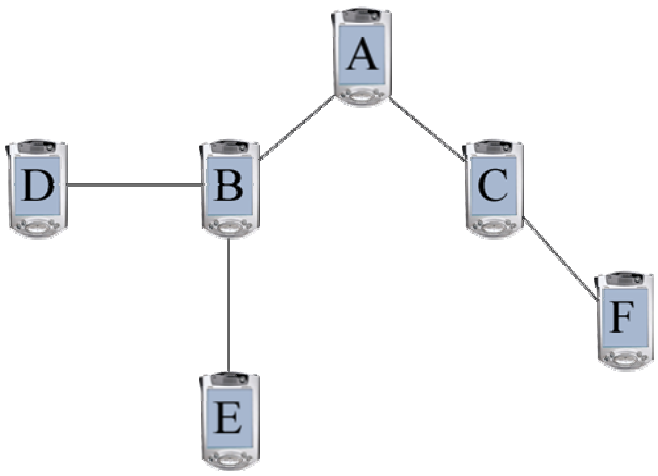

Figure 10. The scenario of the relative geographical position of the nodes.

Figure 11 shows an example of how the internal routing table may look like when the optional multiple path feature is used.

Destination	Gateway	State	Time
Node B	-	active	2004-04-01 12:37:12
Node C	-	active	2004-04-01 12:37:14
Node C	Node B	inactive	2004-04-01 12:37:15
Node B	Node C	inactive	2004-04-01 12:37:15
Node D	Node B	active	2004-04-01 12:37:16
Node E	Node B	active	2004-04-01 12:37:18
Node F	Node C	active	2004-04-01 12:38:05

Figure 11. Example of partial routing table information using the optional multiple path feature from node A's point of view.

5.3.2 Pending List

Due to the limited amount of space in an ARP message, a rebroadcasted ARP request can not contain information about the originator. Therefore, a pending list is introduced. The purpose of the pending list is to maintain information about sent and received ARP requests, replies and forwards.

To prevent network flooding of ARP messages, rebroadcasting of ARP requests destined for a specific destination is only permitted once within an adjustable time limit since the last sent request for the specific destination.

Received ARP requests in the pending list are stored for an adjustable time before they are removed.

Figure 12. Three nodes in an ad hoc network.

Figure 12 shows three nodes in an ad hoc network where each node only can communicate with adjacent nodes. In the case where node A wants to communicate with node C, an ARP request is broadcasted. Without any ad hoc routing protocol present, communication would not be established. A flowchart of how MEARP establishes communication follows:

- Node A broadcasts an ARP request in order to communicate with node C. The ARP request is transmitted autonomous by the operation system.
- Node B receives the incoming ARP request and verifies that it can rebroadcast the message. If node B has transmitted an ARP request, intended for node C within the adjustable time as mentioned earlier, node B ignores the incoming message.
- Node A and node C receives the ARP request sent by node B.
- Node A ignores the message, since its pending list will not permit the message to be rebroadcasted.
- Node C realizes that the ARP request is intended for itself. MEARP therefore ignores the message, whereas the operation system transmits an ARP reply to node B.

- Node B receives the ARP reply sent by node C. Node B now checks with the pending list if any node has requested a route towards node C, and thereafter transmits an ARP forward message to all nodes that has requested a route for node C. In the case described, node B transmits an ARP forward message to node A. Thereafter, node B sets the entry in the pending list as delivered.
- Node A receives the ARP forward message and has thereby a valid route towards node C. Node A also verifies that no other node requests a path to node C. In this case, node A realizes that node B have in fact requested a route, however this is ignore due to the fact that node B gave node A the route.

Figures 13 to figure 17 show each nodes route table and pending list in this example. Note that node C's pending list is empty due to the fact that this node has not transmitted any ARP Request.

Destination	Gateway	State	Time
Node B	-	active	2004-04-01 12:37:12
Node C	Node B	active	2004-04-01 12:37:14

Figure 13. Node A's route table.

Destination	Originator	Delivered	Time
Node C	Node A	yes	2004-04-01 12:37:12
Node C	Node B	no	2004-04-01 12:37:14

Figure 14. Node A's pending list.

Destination	Gateway	State	Time
Node A	-	active	2004-04-01 12:37:12
Node C	-	active	2004-04-01 12:37:14

Figure 15. Node B's route table.

Destination	Originator	Delivered	Time
Node C	Node A	yes	2004-04-01 12:37:12

Figure 16. Node B's pending list.

Destination	Gateway	State	Time
Node B	-	active	2004-04-01 12:37:12

Figure 17. Node C's route table.

5.3.3 Link Quality Aspects

MEARP is a redundant ad hoc network protocol, as mentioned in chapter 5.3.1; MEARP could be configured to manage multiple paths in the routing table. To be able to keep a route up to date and active, cross-layer issues are taken under consideration. There are several ways to monitor and determine the route performance of an active route.

By processing these issues, a metric is created for all active routes. This metric gives a value of the performance of the route. Considering this metric a new route towards the destination can be created in advance.

This chapter handles some of the cross-layer issues that can help to maintain and renew a route.

5.3.4 Route Stability Issues

Every wireless communicating device has limitations in radio transmitting range due to limited signal power, environmental interference etc.

In ad hoc networks the main problem is to provide stable links between communicating devices and still reduce the number of intermediate nodes, without affecting the link quality.

In figure 18, the most efficient way for node A to communicate with node C is to use node B as an intermediate node. That is, the link between node A and node C would probably work; however, great packet loss will occur due to the limited range of node A. By using node B as a relay node, a more stable link between node A and node C is established.

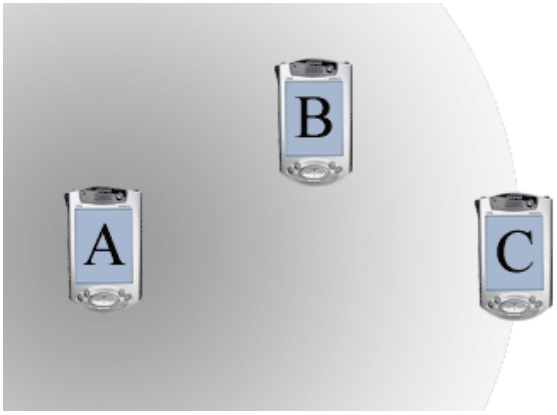

Figure 18. Stable link provided by using intermediate node as relay station.

5.3.5 *Data Link Layer Feedback*

To provide stable links between nodes, data link layer feedback can be used. The general idea is to make use of existing information in the data link layer, used to ensure data link communication. The IEEE 802.11 protocol supports a feature called MAC-layer acknowledgement, which verifies for the transmitter that the receiver de facto has received the message [13].

Therefore, usage of MAC-layer acknowledgement provides information for MEARP when a link is broken or about to be broken. However, consideration must be taken that if the link is about to be broken and this is the only available link to a certain destination, the link should be used rather than be dismissed.

Further, information such as error detection and correction from the Logical Link Control (LLC), see chapter 3.2.2.1, could be used to foresee broken links and thereby finding alternative routes in advance.

5.3.6 Transport Layer Feedback

To provide even more stable links, transport layer feedback can be used as a complement to the data link feedback. The transport layer feedback applies only to protocols with resending capabilities such as the Transmission Control Protocol (TCP). By using information about erroneous messages, routes could be switched in advance to prevent link failures and thereby providing stable routes. TCP, for example, demands packet acknowledgement. If no acknowledgement has been received the packet is resent. The transport layer feedback is thereby provided by tracing such information. Further, bit error counting could provide the information necessary to dismiss a route.

5.3.7 Signal to Noise Ratio

The Signal to Noise Ratio (SNR) is defined as the ratio of the measured signal power to the power of the error. The SNR is measured in dB, and is given by the formula:

$$SNR = 20 \cdot \log_{10}\left(\frac{P_s}{P_e}\right)$$

Where P_s is the signal power and P_e is the power of error.

Ideally the signal power is much greater than the noise level, and the data communication is reliable. If the signal is weak but still above the noise level, the data communication will suffer from reduction of data speed and packet loss. When the signal power is less than the noise level the SNR will be negative and reliable communication will generally not be possible.

When node topology changes in the ad hoc network the SNR level will change. The level can also change by other circumstances i.e. low battery power. By monitoring the SNR level the route loss to a node can be predicted and a new route can be established in advance.

5.4 Simulations

Currently simulations are taking place in a network simulation tool called OPNET. However due to lack of time; the final test results are not finished at this time.

5.5 The Implementation

To ensure that the theories about MEARP work in the reality, we made an implementation. The purpose has been to provide a testbed as well as experience for future development, rather than an optimal solution.

Realization of MEARP has been made in a Red Hat 9, kernel 2.4.22; a Linux based operating system, on laptop computers equipped with Wireless Local Area Network (WLAN) IEEE 802.11b cards. The choice of operating system was based on earlier work we have done within similar field of subject. One of the benefits of using Linux as development platform is that it is an open source system. This provides the possibility to modify the behavior of drivers for network interfaces, protocol stacks etc.

The IEEE802.11b standard is not the optimal communication strategy to use for ad hoc networks; however, it is the most commonly used and easiest to demonstrate the ad hoc functionality with.

MEARP has been implemented as a passive routing protocol, that is, without interfering, interrupting or preventing communication in progress. All source code is developed in the programming language C.

The source is implemented in the Linux user-space as a program and not as a part of the kernel.

During development and testing we have used a program called MacKill [14], a filter tool, used to simulate connectivity configurations. MacKill is primarily used to filter data packets received from a specific source.

To provide link quality, see chapter 5.3.3, signal to noise ratio (SNR) is measured at the receiver. A mean value of the SNR is then transmitted to the transmitter within an adjustable time limit. If no information about the SNR is received by the transmitter, a new route may be requested and chosen. However, a more reliable and optimal way of ensuring link quality would be to use data link layer feedback as described in chapter 5.3.5, combined with transport layer feedback, chapter 5.3.6.

Routes and choice of active routes are maintained by the internal routing table in MEARP, whereas the actual IP routing is performed by the Linux kernel.

Intelligence to discover inconsistencies in routes has been implemented to provide loop free routes. For instance, node A has a route to node C through node B. Node B on the other hand has a route to node C through node A. As described both node A and node B has routes through each other to reach node C. Without any intelligence to remove inconsistent routes, communication with node C is not possible. However, by allowing nodes to investigate from who they received a packet and whereto they are supposed to transmit the same packet, nodes may realize that they are the same and therefore remove the route.

5.5.1 Internet Gateway Support

To be able to access the ad hoc network from the outside or for the ad hoc network to access other networks such as Internet, a gateway support is enabled in the implementation.

To allow the gateway support, Network Address Translation (NAT) [15] is applied. A NAT gateway is used to hide the networks internal addresses towards the outside network. All packets of a TCP or UDP session pass through the same NAT. The gateway node uses two Ethernet interfaces, one wireless interface for communication to the ad hoc network, and an interface for external communication.

When a packet is found to be for an external host, it is routed to the nodes default gateway. The packet is forwarded to the gateway without changing the destination address. When the packet reaches the networks gateway node the initial packet is routed towards the final destination in the global Internet.

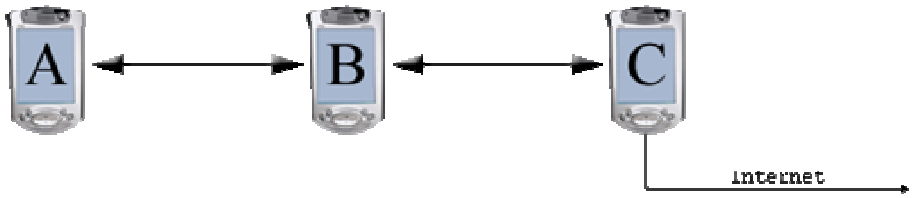

Figure 19. Gateway communication where node C acts as gateway to the Internet.

In figure 19, node A makes a request for an external server on the Internet. The packet is sent on to node A's default gateway, in this case node B. Node B receives the packet and the IP forwarding mechanism forwards the packet towards node C. Node C receives the packet and forwards it on to its external interface, and the packet is routed towards the final destination.

5.5.2 Initial Test Results

So far, MEARP has been tested in an eight node environment. The tests have been carried out by using different transfer protocols such as UDP, used primarily for streaming audio and video, and TCP, used for testing the Internet gateway support.

The video streaming has been performed over seven transmissions, that is, seven hops. The protocol has proven itself to be more stable and efficient than we could have imagined. However, optimization of the implementation is needed to provide even less delays and time for rerouting. Also, the Internet gateway support has proven itself to be successful.

During the tests, which are not included in this report, we dynamically changed the network topology and even removed nodes.

Further, comparison has been made with another ad hoc routing protocol. For this purpose we choose the AODV protocol and an implementation by the University of Uppsala [16] (Version 0.7.2), which can be seen as the most recognized implementation as of today.

The comparison has shown that MEARP is a sufficient competitor with great possibilities. However, both protocols require code optimization before a scientific comparison can be made.

5.5.3 Further Development

The work and development of MEARP has proven the strategy of using and reusing ARP messages for ad hoc purpose to be successful. The current realization of MEARP is fully functional and serves its purpose as expected. During the development we encountered some issues to improve the performance of MEARP.

The program execution rate can be enhanced by code optimization. By considering the information provided by different OSI layers traffic control can be gained. This gives us more detailed information about the routes, and route management is to be more accurate and predicting, which will speed up the route finding procedure.

The current version of MEARP supports only one gateway, and this gateway must be configured when starting up at all nodes. This can be improved with an automatic Internet gateway discovery mechanism. To make the gateway discovery even more efficient we suggest support for multiple gateways. By using multiple gateways, such as Bluetooth™, WLAN, GPRS, and wired connection there will almost certainly be a connection available for communication towards the outside of the ad hoc network.

As mentioned earlier in this chapter MEARP is a reactive, on-demand routing algorithm. The protocol is best suited for slow moving and static environments. By making MEARP experienced based, aware of the movements, routes will be maintained and supervised in a more intelligent way.

The implementation is developed in a Linux environment. By making an implementation in a Windows environment more people will be able to benefit from MEARP.

Further, the optimal implementation to reach the market is to provide the protocol embedded in a radio device. This provides less system resource utilization and less administration of the operating system.

Chapter 6

Conclusions

Ad hoc routing is a growing technology, the market for wireless communication is expanding and the demand for decentralized applications is constantly increasing. E.g. you might wish to download traffic information from your vehicle. To ensure the access between wireless devices without any existing infrastructure, such as cellular phones and wireless computers, ad hoc routing can be a low-priced and effective solution.

During our research and the development of a new ad hoc routing strategy, we came across the Address Resolution Protocol (ARP). The ARP protocol is used to maintain a correlation between the Media Access Control (MAC) address and the Internet Protocol (IP) address, used in network communication. In every IP version 4 (IPv4) network of today the ARP protocol is used to initiate communication.

The main goal for our research has been to develop a new routing strategy, using a minimal of overhead data traffic used to maintain and establish routes in mobile ad hoc routing networks. Many of the existing strategies of today are wasting these resources. By extending the ARP protocol with two new operation types; ARP Route Forward to allow intermediate nodes to send ARP Replies, and ARP Error to allow intermediate node to tell other nodes that a connection is lost, we have successfully managed to create a low data overhead routing strategy, thanks to the ARP Request, and ARP Reply messages already existing and used in the IPv4 communication. This strategy has been named Multi-hop Enabled ARP (MEARP) and is compatible to common IPv4 nodes, thanks to keeping the ARP protocols ordinary functionality.

Further, a demand was to have the protocol taking the link quality in consideration when operating routes. In this way undesirable routes can be avoided. It may be better to use an extra hop to avoid packet loss.

The result of our work has successfully proven that the ARP protocol can be used to support multi-hop data communication.

The conclusion we have reached is that by using existing protocols already used in network environments, you will minimize the data overhead traffic. As a substitute of using control messages to acknowledge routes, link quality and cross-layer information should be taken in consideration, e.g. Transmission Control Protocol (TCP) and MAC layer information.

Chapter 7

Abbreviations

AODV..... Ad hoc on-demand distance vector
ARP ..... Address resolution protocol
DSR ..... Dynamic source routing
FT..... Full topology
GPRS ..... General Packet Radio Services
GPS..... Global positioning system
IEEE..... Institute of electrical and electronics engineers
IETF..... Internet engineering task force
IP..... Internet protocol
IPv4..... Internet protocol version 4
IPX..... Internetwork packet exchange
ISO..... International organization for standardization
LAN ..... Local area network
LLC..... Logical link control
MAC ..... Media access control
MANET..... Mobile ad hoc networks
MEARP ..... Multi-hop enabled ARP
MPR..... Multipoint relays
NAT ..... Network address translation
OLSR ..... Optimized link state routing
OSI..... Open system interconnection
PT..... Partial topology
RERR..... Route error
RF ..... Radio frequency
RFC..... Request for comments
RREP ..... Route reply

RREQ..... Route request
SNR..... Signal to noise ratio
TBRPF ..... Topology dissemination based on reverse-path forwarding
TCP ..... Transmission control protocol
TP..... Topology control message
UDP ..... User datagram protocol
WLAN ..... Wireless LAN
ZRP ..... Zone routing protocol

Chapter 8

References

[1] Ad Hoc Networks with Unattended Ground Sensors
B. Karlsson, A. Lundström, M. Westbergh
Thesis for bachelor of science degree in Telecommunication
University of Kalmar 2002

[2] Ad Hoc Networking
C. Perkins
Addison-Wesley – December 2000
ISBN: 0-201-30976-9

[3] Mobile Ad-hoc Networks (MANET) Charter
IETF MANET Working Group
May 2004
http://www.ietf.org/html.charters/manet-charter.html

[4] Sensor Networks for Network-Centric Warfare
Planning Systems Inc.
October 2000
http://www.plansys.com/Content/NavigationMenu/Products/Sensor_Network_and_Data_Acquisition_Products_White_Papers/Sensor_Networks_for_Network_Centric_Warfare_NCW00.pdf

[5] Topology Dissemination Based on Reverse-Path Forwarding (TBRPF)
R. Ogier, F. Templin, M. Lewis
February 2004
http://www.ietf.org/rfc/rfc3684.txt

[6] Optimized Link State Routing Protocol (OLSR)
T. Clausen, P. Jacquet
October 2003
http://www.ietf.org/rfc/rfc3626.txt

[7] Ad hoc On-Demand Distance Vector (AODV) Routing
C. Perkins, E. Belding-Royer, S. Das
July 2003
http://www.ietf.org/rfc/rfc3561.txt

[8] The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks (DSR)
David B. Johnson, David A. Maltz, Yih-Chun Hu
April 2003
http://www.ietf.org/internet-drafts/draft-ietf-manet-dsr-09.txt

[9] Zone Routing Protocol (ZRP)
N. Beijar
April 2004
http://www.tct.hut.fi/opetus/s38030/k02/Papers/08-Nicklas.pdf

[10] Internetworking Basics
Cisco Systems Inc.
February 2002
http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito_doc/introint.pdf

[11] An Ethernet Address Resolution Protocol
D. Plummer
November 1982
http://www.ietf.org/rfc/rfc0826.txt

[12] Guide to IP Layer Network Administration with Linux
M. Brown
April 2003
http://linux-ip.net/html/

[13] Understanding Link Quality in 802.11 Mobile Ad Hoc Networks
G. Gaertner, V. Cahill
January 2004
http://dsonline.computer.org/0401/f/w1spotp.htm

[14] MacKill v. 0.1
E. Nordström, Uppsala University
February 2004
http://user.it.uu.se/~henrikl/aodv/

[15] The IP Network Address Translator (NAT)
K. Egevang, P. Francis
May 1994
http://www.ietf.org/rfc/rfc1631.txt

[16] Ad-hoc On-demand Distance Vector Routing (AODV-UU)
E. Nordström, Uppsala University
February 2004
http://user.it.uu.se/~henrikl/aodv/