

UNIVERSITÉ LIBRE DE BRUXELLES
Faculté des Sciences
Département d'Informatique

Routing protocols for interconnecting cellular and ad-hoc networks

Mémoire présenté en vue de l'obtention
du grade de Licencié en Informatique

Bayani Carbone
Année académique 2005–2006

Abstract

This study presents an original routing protocol for interconnecting and extending cellular networks, in particular UMTS networks, with Ad-Hoc networks. This new protocol GWAODV¹ is based on already established Ad-Hoc routing protocols: AODV[1] and ABR[2]. Simulations were run using the NS-2[3] simulator to analyze the protocol's performance in terms of end-to-end delay, packet delivery fraction and routing overhead.

¹Gateway Ad-hoc On Demand Distance Vector

Acknowledgments

Above all, I would like to thank my parents and my girlfriend, without whom I would never have been able to finish this work. They have supported me since day one and I am very grateful to have them in my life.

I would also like to thank my family and friends for their never ending support and their precious advises.

Of course, none of this would have been possible without Mr J-M. Dricot, Professor G. Bocq and Professor R. Devillers who found this subject for me and followed me every step of the way. For this I thank them and it has been a pleasure working with them.

Contents

Abstract	2
Acknowledgments	3
1 Introduction	9
I State Of The Art	11
2 Cellular Networks	12
2.1 Introduction	12
2.2 UMTS services	13
2.3 UMTS Architecture	13
2.3.1 Core Network	15
2.3.2 UMTS Terrestrial Radio Access Network	15
2.3.3 User Equipment	18
2.4 Cellular Networks Limitations	19
2.5 3GPP System-WLAN Interworking	19
2.5.1 3GPP-WLAN Interworking Scenarios	20
3 Mobile Ad-Hoc Network (MANET)	22
3.1 Introduction	22
3.2 Ad-Hoc Routing	23
4 Interconnecting Ad-Hoc and Cellular Networks	30
4.1 Existing Solutions	30
4.1.1 iCAR	30
4.1.2 UCAN	33
4.1.3 Conclusion	34
4.2 Motivations of this study	34
4.2.1 Propagation Issues	34

II	Personal Contribution	42
5	AODV: Ad-hoc On-demand Distance Vector	44
5.1	Message Types	44
5.2	AODV Operation	45
5.2.1	Sequence Numbers	45
5.2.2	Routing Table	45
5.2.3	Route Discovery	46
5.2.4	Route Repair and RERR	49
6	ABR: Associativity-Based Routing	51
6.1	ABR Operation	51
6.1.1	Route Discovery	51
6.1.2	Route Repair	52
6.1.3	Route Delete	53
7	Hypotheses and Assumptions	54
7.0.4	User Equipment (UE)	54
7.0.5	UMTS model	54
7.0.6	Traffic Direction	54
7.0.7	Security and Authentication	55
7.0.8	Charging and Billing	55
8	GWAODV: Gateway AODV	56
8.1	Introduction	56
8.2	Hello-Messages	56
8.3	Neighbour Table and Associativity	56
8.4	Metrics	58
8.5	Gateway Table	58
8.6	Gateway Selection Algorithm	59
8.6.1	Algorithm Operation	59
8.7	Creation of Gateway Hello-Messages	60
8.8	Reception of Gateway Hello-Messages	60
8.9	Deleting Gateway Routes	61
8.10	Illustration	62
8.10.1	Hello-Messages	62
8.10.2	Gateway Selection	62
8.11	Data Packet Processing	64
8.12	Gateway Route Repair	65
9	NS-2: Network Simulator	66
9.1	Introduction	66
9.2	Simulator Design	66
9.2.1	Network Components	67

CONTENTS

9.2.2	Event Scheduler	68
9.2.3	Wireless Node	68
9.2.4	Packet	69
9.2.5	Agent	72
10	Simulations	73
10.1	Configuration	73
10.2	Modifications to NS-2	74
10.2.1	Double Interface Node	74
10.2.2	Modifications to AODV	74
10.2.3	Modifications to UMTS_MAC_FDD	76
10.2.4	Modifications to NOAH	76
10.2.5	Other Modifications	76
10.3	Simulation Configurations	76
10.4	General Performance	77
10.4.1	Results	78
10.5	Different Propagation Models	81
10.6	Varying Packet Rate	84
10.7	Conclusion	84
11	Conclusion and Future Work	86
	Bibliography	88

List of Figures

2.1	Example of UMTS network implementation	16
2.2	Scenarios and their Capabilities[7]	21
3.1	Operation modes of IEEE 802.11	23
3.2	Example of wireless mesh network	24
3.3	Partial Ad-Hoc Routing Protocols Classification	25
4.1	Primary Relaying	31
4.2	Case one of Secondary Relaying	31
4.3	Case two of Secondary Relaying	32
4.4	Cascaded Relaying	32
4.5	Example of UCAN use	33
4.6	Average power received when transmitting one Watt of power using the free space model	35
4.7	Example of multipath signal	37
4.8	Cellular network assisted by Ad-Hoc networking	41
5.1	Example of the need of a route discovery phase	46
5.2	RREQ creation and distribution	47
5.3	RREQ forwarding and RREP creation	49
6.1	ABR Route Discovery	52
8.1	Increasing Associativity Level	57
8.2	Decreasing Associativity Level	58
8.3	Routing loop problem	61
8.4	First Hello-Message Exchange	62
8.5	First Hello-Message Received	63
8.6	Gateway Table	63
9.1	Correspondence between the C++ and OTcl hierarchies	67
9.2	Partial class hierarchy	68
9.3	Event Scheduler	69
9.4	Schematic of a wireless node	70
9.5	Node Configuration	71

LIST OF FIGURES

9.6	Packet Format	71
10.1	Schematic of a double interface wireless node	75
10.2	Setdest Parameters	78
10.3	Average End-To-End Delay per maximum number of connections	78
10.4	Average End-To-End Delay as a function of time	79
10.5	Average Packet Delivery Fraction per maximum number of connections	80
10.6	Average Packet Delivery Fraction as a function of time	81
10.7	Average Overhead per maximum number of connections	82
10.8	Packet Delivery Fraction per Simulation	82
10.9	Amount of dropped packets due to the lack of gateway routes	83
10.10	Average End-To-End Delay per Simulation	83
10.11	Average End-To-End Delay per Sending Rate	84
10.12	Packet Delivery Fraction per Sending Rate	85

Chapter 1

Introduction

The rapid evolution of wireless technology has made people's needs for communication and access to information grow even faster. The fact that anyone can be reached at any time and any place has become extremely convenient, for some it has even become mandatory. The advent of the Internet has made any information easily accessible and we now expect the same on-the-go.

Wireless communications have been in constant evolution for the past few years, the best known example being cellular networks which provide people with communication services along with the freedom of movement. Unfortunately physical constraints that arise when working with wireless technology make it difficult to provide such services everywhere, especially indoors.

Radio signals are affected by the way buildings are constructed. What materials are used and which indoor layout is chosen can deeply influence indoor radio signal propagation. Sometimes the effect is positive but usually it's the contrary even creating "dead spots" where the signal is completely attenuated. Since tearing down and re-constructing buildings from scratch with radio signal propagation in mind is not a reasonable solution, we have to work within the current environment and find solutions to extend coverage.

This work presents a solution using Ad-Hoc networking when reception from a base-station is impossible. Using other mobile devices as relay points to reach a base-station we can provide access to cellular networks which otherwise would not be within reach. Of course, if a node is completely isolated, has no neighbouring node then it will remain out of coverage.

The first part of this study presents the general concepts of cellular networks and Ad-Hoc networks. Then, existing work and the motivations for interconnecting these network infrastructures will be explained.

CHAPTER 1. INTRODUCTION

The second part of this study is the personal contribution. It starts by a detailed view of two routing protocols whose concepts are used in the solution presented in this work.

Then, follows a complete description of the new routing protocol along with the hypotheses and limitations of this work. The study ends with a description of the network simulator used to analyze the protocol's performance and the results it generated.

Part I
State Of The Art

Chapter 2

Cellular Networks

2.1 Introduction

Cellular networks emerged with the introduction of first-generation wireless telephone technology (1G). Prior to 1G, so-called 0G technologies were used as solutions for mobile telephony from the late 40s to the early 80s. What separated 0G systems from previous wireless communication system was the fact that 0G systems were available as a commercial service that was part of the public switched telephone network rather than part of a closed network such as the taxi dispatch system for instance. 0G technologies usually used a limited number of channel (23 in IMTS: Improved Mobile Telephone System) and a single base station.

In order to respond to network congestion and power consumption issues of 0G, 1G brought a cellular concept to mobile telecommunications. 1G technologies are characterized by analog voice communication although digital signaling was used between the terminals and the base stations. A very popular system used in the USA was AMPS (Advanced Mobile Phone System) developed by Bell Labs and installed in 1982. It was also used in other countries under different names (TACS in England, MCS-L1 in Japan). These systems introduced the concepts of cells and frequency reuse which are found in second-generation wireless telephone systems.

The next step or second-generation wireless telephone technology (2G) made the transition to an all digital system adding digital voice communication to the already present digital signaling. Systems such as GSM and CDMA are still used today, GSM being the most globally widespread system to date.

However with the development of the Internet, the need for data services has grown considerably not only using wired devices like personal computers but also wirelessly through mobile handsets. Some experts even expect that, as with the standard telephone network, data traffic will surpass voice traffic in mobile communications[4]. Since 2G systems were not designed to support

such services, temporary solutions labeled as 2.5G like GPRS ¹ and EDGE ² have been introduced in recent years to try and satisfy the demand of consumers while a new generation of mobile communication technology is put into place.

For the past couple of years, cellular networks have begun their third and latest evolution (3G). Based on previously established CDMA ³ technology labeled as W-CDMA ⁴, this new generation of mobile telecommunications was given the name: UMTS (Universal Mobile Telecommunications System). This latest chapter in cellular networks focuses on providing data services. Data rates vary from 144Kbps for users in moving vehicles (high mobility), 384Kbps for pedestrians in an urban outdoor environment (full mobility) and 2Mbps for stationary indoor users (limited mobility). This technology will offer what xDSL provides to "wired" users, that is an "always on" broadband data connection with the addition of mobility. GPRS and EDGE already offer an always on connection but the maximum data rate achieved amongst these two systems is at best 473.6 Kbit/s.

2.2 UMTS services

The services offered by an UMTS network consist of the same services present in a GSM/GPRS network as well as broadband services and what is called a "Virtual Home Environment" (VHE). VHE allows a roaming user to enjoy the same services, provided to him by his home network, on a partner network.

UMTS supports quality-of-service by using four different classes of services, each defining certain needs in terms of parameters such as delay, jitter and transmission errors. For time dependent applications such as voice or video communications, delay and jitter are important factors to guarantee a level of quality whereas other applications like SMS or E-mail rely essentially on a communication with a low error rate.

Figure 2.1 shows the different classes of services and example applications for each of them along with the parameters which have to be considered.

2.3 UMTS Architecture

An UMTS network is built upon three interacting domains: the User Equipment (UE), the UMTS Terrestrial Radio Access Network (UTRAN) and the Core Network (CN). This section will present these domains, starting with

¹General Packet Radio Service

²Enhanced Data rates for GSM Evolution

³Code Division Multiple Access

⁴Wideband Code Division Multiple Access

Class Of Service	Applications	Parameters		
		Delay	Jitter	Errors
Conversational	<ul style="list-style-type: none"> • Voice(multiple compression modes) • Telephony, VoIP • Visiophony • Elaborate Video Games 	Low transfer delay	Low jitter	Importance depends on type of compression (more important for visiphony than voice at 12.2kbit/s)
Streaming	<ul style="list-style-type: none"> • Web broadcast • Video streaming on-demand 	Transfer delay less significant	Jitter less important (make use of buffering)	idem
Interactive	<ul style="list-style-type: none"> • Database Consultation • Localization services • Simple Video Games 	Transfer delay less significant (based on round trip delay)	Not so important	Important
Background	<ul style="list-style-type: none"> • E-mail • SMS • Measurements 	Not so important	Not so important	Important

Table 2.1: UMTS services[5]

the Core Network.

Figure 2.1 presents an example implementation of an UMTS network.

2.3.1 Core Network

The Core Network for UMTS is based on the existing GSM/GPRS network; however every network equipment needs to be altered so as to support UMTS services. The main purpose of the CN is to supply switching, routing and transit for user traffic. Databases and network management functions are also found in the CN.

The CN itself consists of two domains: the circuit switched domain and the packet switched domain. The Mobile services Switching Centre (MSC), Visitor location register (VLR) and Gateway MSC are located in the circuit switched domain while the Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN) reside in the packet switched domain. Nevertheless some entities like EIR, HLR, VLR and AUC are shared by both domains.

For transmission within the CN, ATM (Asynchronous Transfer Mode) is the chosen technology. Circuit switched traffic is managed by the type 2 AAL (ATM Adaptation Layer) which is designed for delay sensitive applications with variable bit rates. Packet switched traffic on the other hand is handled by the type 5 AAL designed to transport data frames for connectionless data services.

The architecture of the Core Network is not in a frozen state, it might have to be modified in order to support additional services. Furthermore, existing equipments such as the MSC, VLR and SGSN can be merged to become an UMTS MSC.

2.3.2 UMTS Terrestrial Radio Access Network

The UTRAN consists of radio network controllers or RNCs and base stations called NodeBs which are connected to these RNCs. Its goal is to provide connectivity between the User's Equipment (UE) and the Core Network.

As mentioned above, a NodeB uses W-CDMA as air transport technology. The main purpose of NodeB is the conversion to and from the radio interface. A NodeB is connected to a RNC and can serve multiple cells.

Tasks of a NodeB:

- Transmission and reception of data across the radio interface
- Forward error correction (FEC)
- Rate adaptation

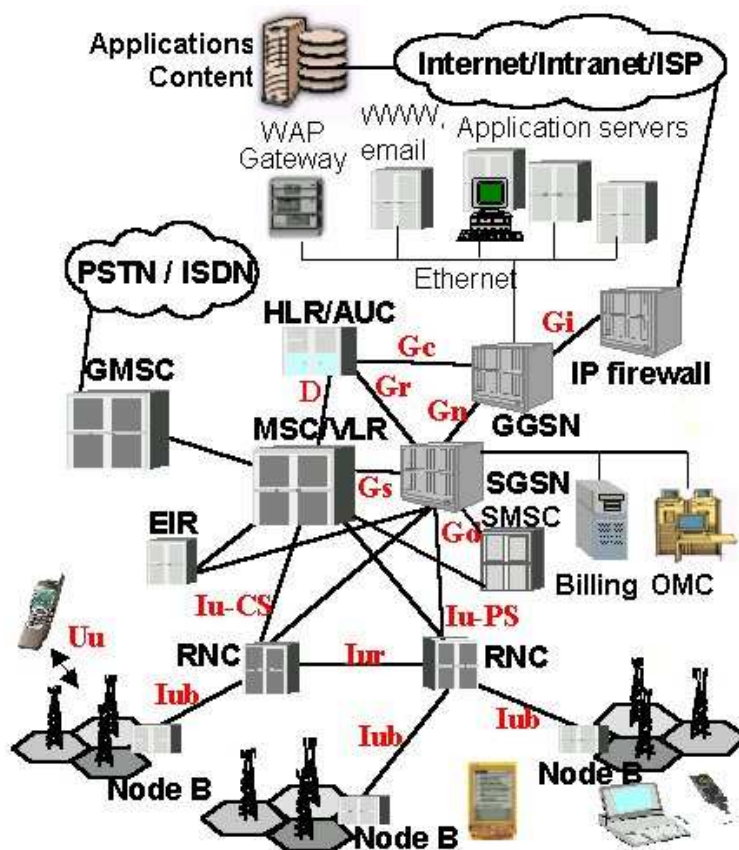


Figure 2.1: Example of UMTS network implementation

- W-CDMA spreading/despreading
- Apply the codes that are necessary to describe channels in a CDMA system
- Quadrature Phase Shift Keying (QPSK) modulation on the air interface
- Power control: through the inner-loop power control, the NodeB commands the UE to adjust its transmission power based on up link transmission power control information (i.e. measurements coming from the UE). The predefined values for inner-loop power control are derived from the RNC via outer-loop power control
- Send measurements report to the RNC for handover and macro diversity combining
- Carry out "Softer Handover" which reduces the amount of transmissions between a NodeB and its RNC and insures micro diversity

A RNC controls one or more NodeBs and enables autonomous radio resource management (RRM) by the UTRAN. A RNC and its associated NodeBs make up the Radio Network Subsystem (RNS)

Tasks of a RNC:

- Radio Resource Control
- Admission Control
- Channel Allocation
- Power Control Settings
- Handover Control, combines data received from different NodeBs for a same UE during "Soft Handover" for instance
- Macro Diversity
- Ciphering
- Segmentation / Reassembly
- Broadcast Signaling
- Outer-Loop Power Control

2.3.3 User Equipment

The UE is the device that subscribers use to access their operator's Core Network and which act as counter parts for NodeBs across the air interface. This device can of course be a mobile handset but also a card in a laptop for example. It is based on the same principles as the GSM Mobile Station specifically there is a separation between mobile equipment (ME) and the UMTS subscriber identity module card (USIM).

The UMTS IC card has the same physical characteristics as GSM SIM card, here are its different functions:

- Support of one or more (optionally) User Service Identity Module (USIM) application
- Support of one or more user profile on the USIM
- Update USIM specific information over the air
- Security functions
- User authentication
- Optional inclusion of payment methods
- Optional secure downloading of new applications

In addition these terminals have multiple identities with most of them already present in the GSM specification.

UE Identities:

- International Mobile Subscriber Identity (IMSI)⁵
- Temporary Mobile Subscriber Identity (TMSI)⁵
- Packet Temporary Mobile Subscriber Identity (P-TMSI)
- Temporary Logical Link Identity (TLLI)
- Mobile station ISDN (MSISDN)⁵
- International Mobile Station Equipment Identity (IMEI)⁵
- International Mobile Station Equipment Identity and Software Number (IMEISV)

Three modes of operation have been identified for a UE:

⁵Identity already present in GSM specification

- PS⁶/CS⁷ mode of operation: The UE is attached to both the PS domain and CS domain, and is capable of simultaneously operating PS services and CS services
- PS mode of operation: The UE is attached to the PS domain only and may only operate services of the PS domain. However, this does not prevent CS-like services to be offered over the PS domain (like VoIP)
- CS mode of operation: The UE is attached to the CS domain only and may only operate services of the CS domain

2.4 Cellular Networks Limitations

3G systems, as well as all the other systems presented above, rely on base stations to gain access to an operator's network. Thus if a mobile handset is out of range of any base station it cannot communicate with other devices using UMTS.

Adding to already present radio propagation issues of other wireless technologies is the cell breath effect observed around NodeBs in UMTS. The cell breath effect underlines the fact that capacity and coverage are not independent in a CDMA environment. When the traffic load of a cell rises, the signal of a UE located at the edge of the cell is submerged by the interference generated by UEs closer to the NodeB thus reducing coverage. If an edge UE is lucky, it will be able to switch to another cell otherwise it is forced to wait for the user to move, or for closer UEs to exit the cell.

This is when other wireless technologies, such as 802.11 in Ad-Hoc or Infrastructure mode, could provide an alternative for "stranded" UEs. In Infrastructure mode the idea would be to allow a UE to use a WLAN access point to access the Internet for instance when reception from a NodeB is not possible. In Ad-Hoc mode, the system presented in this work allows a UE which has no connectivity with a base station to use a series of mobile handsets acting as hops along a route to reach a NodeB and thus the operator's 3G network. These solutions would make it possible for a "stranded" UE to forward its traffic along this alternate route instead of having to wait for 3G network reception to be available.

2.5 3GPP System-WLAN Interworking

Some work has already been done to allow a UE to switch from the UMTS network to a WLAN if needed. The 3GPP⁸ collaboration has issued mul-

⁶Packet Switched

⁷Circuit Switched

⁸3GPP: The 3rd Generation Partnership Project (3GPP) is a collaboration agreement that was established in December 1998. The collaboration agreement brings together

multiple TR's (Technical Report) since 2002, presenting six different scenarios of 3G-WLAN interworking using WLAN as a complementary radio access technology to the 3GPP system.

2.5.1 3GPP-WLAN Interworking Scenarios

Please refer to [7] for a complete description.

Scenario 1: Common Billing and Customer Care

This scenario allows a UE to connect to a WLAN provided by the user's home operator. This enables the user to receive a unique bill including both 3G and WLAN services charges. The user has access to Internet services and resources from the WLAN but does not have access to 3GPP services or resources other than those he can normally access from the Internet. The security level of the two systems may be independent; for instance, the operator can grant a username and password for WLAN access to users who subscribe to such a service.

Scenario 2: 3GPP system based Access Control and Charging

The main difference between this scenario and the previous one is that authentication, authorization and accounting are provided by the 3GPP system which means that the user does not notice a significant difference in the way that access is granted. More importantly this makes it easier for an operator to grant access to the WLAN service for desiring subscribers.

Scenario 3: Access to 3GPP system, PS based services

This scenario allows users to access 3G PS services (MMS for example) through the WLAN. However, service continuity between the 3GPP system and the WLAN is not required.

Scenario 4: Service Continuity

The purpose of this scenario is to allow users to switch from the 3G system to the WLAN system and vice versa without having to reestablish active services. The two systems having quite different characteristics there may be a change in service quality and the user may notice the switch. Some services not supported by the destination system of a switch may however be disconnected.

a number of telecommunications standards bodies which are known as "Organizational Partners". The current Organizational Partners are ARIB, CCSA, ETSI, ATIS, TTA, and TTC. The original scope of 3GPP was to produce globally applicable Technical Specifications and Technical Reports for a 3rd Generation Mobile System based on evolved GSM core networks and the radio access technologies that they support [6]

Scenario 5: Seamless services

What sets this scenario apart from the previous one is the fact that an effort is placed on minimizing aspects such as data loss and break time during the switch between access technologies for services such as the ones supported by Scenario 3: PS based services. An example is maintaining a VoIP session throughout a switch from WLAN to 3G without noticeable interruption.

Scenario 6: Access to 3GPP CS Services

This last scenario is aimed at providing users with 3G CS services through a WLAN system without implying any circuit-switched characteristics in this system. Seamless switching between systems is also part of this scenario.

Figure 2.2 sums up the capabilities of the different scenarios.

Scenarios:	Scenario 1: Common Billing and Customer Care	Scenario 2: 3GPP system based Access Control and Charging	Scenario 3: Access to 3GPP system PS based services	Scenario 4: Service continuity	Scenario 5: Seamless services	Scenario 6: Access to 3GPP system CS based Services
Service and operational Capabilities:						
Common billing	X	X	X	X	X	X
Common customer care	X	X	X	X	X	X
3GPP system based Access Control		X	X	X	X	X
3GPP system based Access Charging		X	X	X	X	X
Access to 3GPP system PS based services from WLAN			X	X	X	X
Service Continuity				X	X	X
Seamless Service Continuity					X	X
Access to 3GPP system CS based Services with seamless mobility						X

Figure 2.2: Scenarios and their Capabilities[7]

Chapter 3

Mobile Ad-Hoc Network (MANET)

3.1 Introduction

A MANET is a network where nodes are both hosts and routers. These nodes are able to move around freely inside the network but also to exit and enter the network at any time. The result of this mobility is that the network topology is constantly changing which makes it impossible to use established fixed routing algorithms "as-is". An important factor in Ad-Hoc networking is power consumption. Since MANETs are usually designed for mobile nodes running on batteries an effort has to be made to preserve battery life especially in routing protocols.

Ad-Hoc networking has become popular in recent years with the arrival of 802.11¹, the technology which provides a low-cost wireless solution. Figure 3.1 shows the two types of operation mode of IEEE 802.11. However, MANETs are not bound to 802.11; in fact these networks have been around since even before the Internet when they were called "packet radio" networks and sponsored by DARPA² in the early 70s.

As with many communication technologies, Ad-Hoc networking was first developed for military use such as the Joint Tactical Radio System (JTRS) that the U.S. Army uses in field operations. Nevertheless, the area of application of these networks is rather vast and includes situations like a fleet of ships at sea where no infrastructure is present, emergency workers in disaster zones where the infrastructure has been destroyed and more commonly when a group of people want to exchange information without having to set-up an infrastructure if none is present or simply to avoid using any infrastructure.

¹IEEE 802.11 a set of Wireless LAN/WLAN standards developed by working group 11 of the IEEE LAN/MAN Standards Committee (IEEE 802)[8]

²Defense Advanced Research Projects Agency

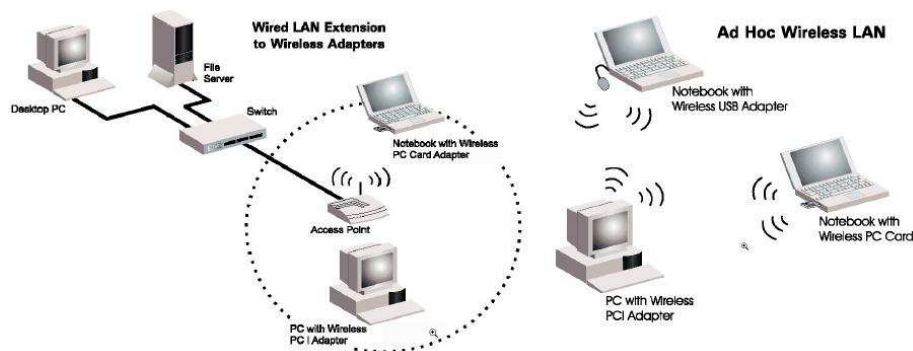


Figure 3.1: Operation modes of IEEE 802.11

Wireless mesh networks is a category of networks immediately derived from MANETs. In these types of networks, nodes cooperate to relay a message to its destination. Each node has a set of neighbors which it can use to forward messages; this way, if a neighbor disappears (i.e. node movement, node switched off) the node can just pick another neighbor and continue to have access to the network. This feature makes wireless mesh networks extremely reliable and scalable. What's more, a way to increase reliability is simply to add nodes which increases the number of neighbors and therefore means more possibilities for a node to forward messages. Figure 3.2 gives an example of a wireless mesh network.

As seen in figure 3.2, one of the nodes of the network can be a base station or another device connected to the Internet for instance. This allows nodes to send packets over the Internet even though they are not within reach of the base station's signal. This characteristic is the reason why this work puts forward a solution that relies on Ad-Hoc networking.

3.2 Ad-Hoc Routing³

As mentioned above, wired networks routing algorithms cannot be used "as-is" in wireless Ad-Hoc networks due to the mobility of the nodes. The basic routing principals such as Distance-Vector and Link-State have to be adapted to Ad-Hoc networking by taking into account that neighbors could appear and disappear at any moment. This section will present different routing algorithms used within Ad-Hoc networks, including AODV⁴ which is used in this work.

³Protocol and algorithm will have the same meaning in this section

⁴Ad-Hoc On Demand Distance Vector

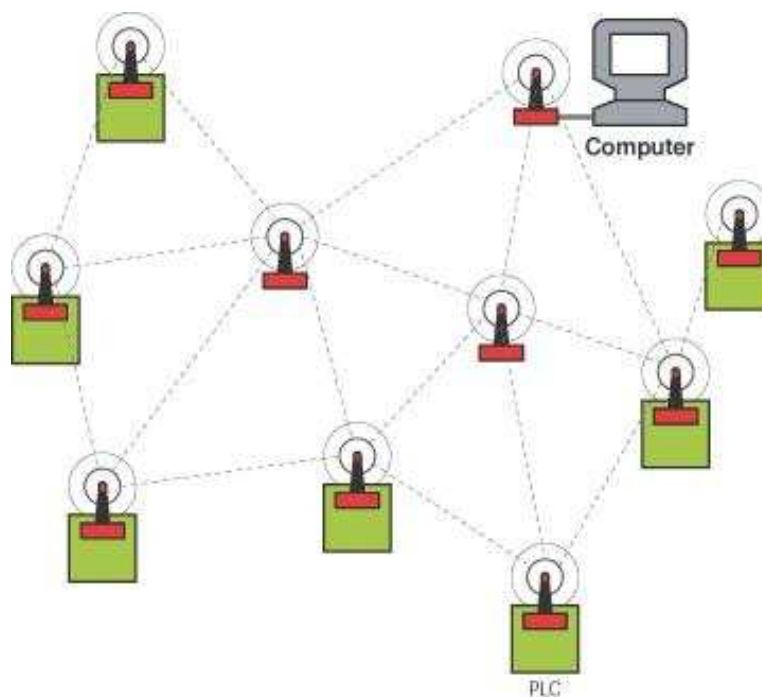


Figure 3.2: Example of wireless mesh network

Ad-Hoc Routing Protocols

Ad-Hoc routing protocols are divided into three classes, which are presented below; figure 3.3 gives a partial classification of Ad-Hoc routing protocols.

1. Proactive Routing Protocols:

These protocols will continuously try to determine the layout of the network. By regularly exchanging packets containing topology information between nodes, a complete picture of the network is maintained at every single node. As a result, the delay in determining the route to be taken is minimal. This is especially important for time-critical traffic.

However, a drawback of proactive protocols is that, due to node mobility, routing information in the tables becomes quickly invalid. Therefore, there are many short-lived routes that are being determined and not used before they disappear. Another drawback resulting from node mobility is the amount of traffic overhead generated when evaluating these unnecessary routes. This is especially aggravated when the network size increases.

Finally, if the nodes transmit infrequently, most of the routing information exchanged becomes redundant. The nodes, nevertheless, continue



Figure 3.3: Partial Ad-Hoc Routing Protocols Classification

to use up energy by continually updating these unused entries in their routing tables which diminishes battery autonomy.

Hence, proactive protocols work best in networks that have low node mobility or where the nodes transmit data frequently or when power consumption is not an issue.

Examples of proactive protocols are:

DSDV[9], Dynamic Destination-Sequenced Distance-Vector Routing. This protocol is based on the classical Bellman-Ford routing algorithm. Each node maintains a list of all destinations and number of hops to each destination. Each entry is marked with a sequence number. It uses full dump (whole routing table) or incremental packets to reduce network traffic generated by route updates. The broadcast of route update is delayed by settling time. The Distributed Bellman-Ford has the looping and count-to-infinity problem, which is avoided in DSDV by using sequence numbers.

DSDV requires a full dump update periodically, therefore DSDV is not efficient in route updating. DSDV limits the number of nodes that can join the network. Whenever the topology of a network changes, DSDV is unstable until update packets propagate through the network. DSDV is effective for creating ad-hoc networks for small populations of mobile nodes.

DSDV is a well-known routing algorithm for ad-hoc network routing.

Because there are no standard specifications, no commercial implementations are available. Many improved protocols based on DSDV have been developed such as AODV.

Another example of proactive routing protocol is OLSR[10] which stands for Optimized Link State Routing Protocol. In this protocol, every node periodically sends broadcast "Hello" messages with information to specific nodes in the network to exchange neighborhood information. Included in these "Hello" messages, are the nodes IP, sequence number and a list of the distance information of the nodes neighbors. After receiving this information a node builds itself a routing table. The node can then calculate the route to every node using the shortest path algorithm. When a node receives an information packet with the same sequence number twice it is discarded. The information in the routing table is updated when a change in the neighborhood is detected, or a route to any destination is expired or when a better (shorter) route is detected for a destination.

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 a classical flooding mechanism, where every node retransmits each message when it receives the first copy of the message. In OLSR, link state information is generated only by nodes elected as MPRs. Thus, a second optimization is achieved by minimizing the number of control messages flooded in the network. As a third optimization, an MPR node may chose to report only links between itself and its MPR selectors. Hence, contrary to the classic link state algorithm, partial link state information is distributed in the network. This information is then used for route calculation. OLSR provides optimal routes (in terms of number of hops). The protocol is particularly suitable for large and dense networks as the technique of MPRs works well in this context.

2. Reactive Routing Protocols:

These kind of protocols only find a route to the destination node when there is a need to send data. The source node will start a route discovery procedure by transmitting route requests throughout the network. The sender will then wait for a response from the destination node or an intermediate node (that has a route to the destination) which will include a list of intermediate nodes between the source and destination.

The main drawback is that route discovery generates a significant delay before the packet can be transmitted. It also requires the transmission of a significant amount of control traffic.

Hence, reactive protocols are most suited for networks with high node mobility or where the nodes transmit data infrequently.

An example of such a protocol is DSR[11] or Dynamic Source Routing. This protocol is based on the concept of source routing. Mobile nodes are required to maintain route caches that contain the routes known by the node. Entries in the route cache are continually updated as new routes are learned. The protocol is composed of the two main mechanisms: "Route Discovery" and "Route Maintenance", which work together to allow nodes to discover and maintain routes to arbitrary destinations in the ad hoc network. To limit the bandwidth used, the process to find a route to a destination is only executed when needed. All aspects of the protocol operate entirely on-demand which implies that DSR is beacon-less, unlike some other reactive protocols such as ABR⁵ for instance. There are no "Hello" messages used between nodes to notify their neighbors about their presence. In DSR the sender (source, initiator) determines the whole path from the source to the destination node (Source-Routing) and inserts the addresses of the intermediate nodes used along the route in the packets it sends.

The protocol allows multiple routes to any destination and allows each sender to select and control the routes used in routing its packets, for example for use in load balancing or for increased robustness. Other advantages of the DSR protocol include easily guaranteed loop-free routing, operation in networks containing unidirectional links and very rapid recovery when routes in the network change.

DSR also suffers from a scalability problem due to the nature of source routing. As the network becomes larger, the control packets and message packets also become larger. This gives a negative impact due to limited bandwidth. Also packets may be forwarded along stale cached routes.

Hence, DSR works best in networks with a small diameter (between 5 and 10 hops) and the nodes should only move around at a moderate speed.

TORA[12], Temporally-Ordered Routing Algorithm is another example of reactive routing protocol. This protocol is a highly adaptive, loop-free, distributed routing protocol based on the concept of link reversal. It is source initiated and provides multiple routes for any desired source/destination pair. There are 3 basic functions of the protocol: route creation, route maintenance and route erasure.

Since this protocol uses internodal co-ordination it exhibits instability behavior similar to "count-to-infinity" problem in distance vector

⁵Associativity Based Routing

routing protocols. There is a potential for oscillations to occur, especially when multiple sets of coordinating nodes are concurrently detecting partitions, erasing routes, and building new routes based on each other. Though, such oscillations are temporary and route convergence will ultimately occur.

Finally, AODV⁶[1]: Ad Hoc On-Demand Distance Vector Routing Protocol is one of the most used Ad-Hoc routing protocol. It is a reactive routing protocol based on DSDV. AODV is designed for networks with tens to thousands of mobile nodes. One feature of AODV is the use of sequence numbers in order to insure loop freedom. Sequence numbers are used by other nodes to determine the freshness of routing information. If a node has the choice between two routes to a destination, a node is required to select the one with the greatest sequence number.

In AODV, every node has a routing table. As with DSR, the process to find a route to a destination is only executed when needed using RREQs (Route Request) messages originating from the source node and RREPs (Route Reply) messages sent by the destination or intermediate nodes which have a route to the destination. AODV also implements a "Route Repair" mechanism which enables a node to switch to a different route if an intermediate node along the original route goes down.

3. Hybrid Routing Protocols:

Since proactive and reactive routing protocols each work best in oppositely different scenarios, there is good reason to develop hybrid routing protocols, which use a mix of both proactive and reactive techniques protocols. These hybrid protocols can be used to find a balance between the proactive and reactive protocols.

The basic idea of hybrid routing protocols is to use proactive routing mechanisms in some areas of the network at certain times and reactive routing for the rest of the network. The proactive operations are restricted to a small domain, called the zone radius, in order to reduce the control overheads and delays. The reactive routing protocols are used for locating nodes outside this domain, as this is more bandwidth-efficient in a constantly changing network.

The perfect example of a hybrid routing protocol is ZRP[13]: Zone Routing Protocol. In this protocol, the radius of each node's local routing zone plays an important part in determining the proactive zone. The proactive mechanism is used to determine the topology

⁶This is a brief summary of this protocol, it will be more widely discussed in the second part of this work (Personal Contribution).

within the radius of the node. The reactive routing protocol is then used to locate nodes outside the radius of the node on demand.

The adjustment of the zone radius will allow the protocol to adapt to different network environments. A larger radius will favor the proactive routing protocol, optimal for slow-moving nodes or large amounts of traffic. Consequently, a smaller zone radius will favor the reactive protocol, which is optimal for fast-moving nodes or small amounts of traffic.

Chapter 4

Interconnecting Ad-Hoc and Cellular Networks

4.1 Existing Solutions

Two existing systems will be presented in this section to illustrate the benefits of interconnecting Ad-Hoc and Cellular networks:

- iCAR[21] which addresses the congestion problem due to unbalanced traffic in a cellular system and provides interoperability for heterogeneous networks and
- UCAN[22], an architecture designed to enhance cell throughput, while maintaining fairness¹.

4.1.1 iCAR

iCAR or Integrated Cellular and Ad hoc Relaying System is a relatively new solution introduced in 2002 by Hongyi Wu.

This system makes use of Ad-Hoc Relay Stations (ARS) which are lightweight² versions of a cellular network's base station. These relay stations are capable of communicating directly with a cellular network base station, mobile hosts and other relay stations. The difference between standard base station and an ARS is that this relay station is a wireless device whereas a base station is usually connected to a controlling equipment (i.e. a BTS connected to a MSC in GSM) using wires. In addition ARS's are capable of limited mobility which allows an operator to move its ARS to strategic areas in order to adapt to changing situations.

¹Providing service to low data-rate users is required for maintaining fairness, but at the cost of reducing the cell's aggregate throughput

²literally and in terms of complexity

Three main relaying situations involve these relay stations and mobile hosts:

- **Primary Relaying:** If a mobile host is unable to obtain a channel in a congested cell, it can be relayed to a neighboring cell using an ARS if the mobile host is within the coverage area of an ARS. This case is illustrated by figure 4.1.

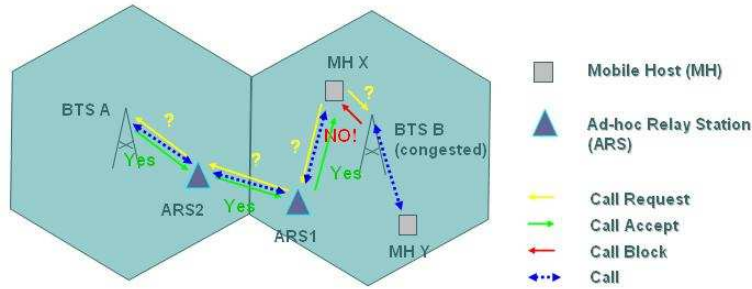


Figure 4.1: Primary Relaying

- **Secondary Relaying:** In this type of relaying, the objective is to free a channel of a congested base station so it can be attributed to a mobile host located outside the coverage area of any ARS. Two cases are possible:
 - Either an on-going connection by a mobile host in the vicinity of an ARS is transferred to another cell which frees up a channel.

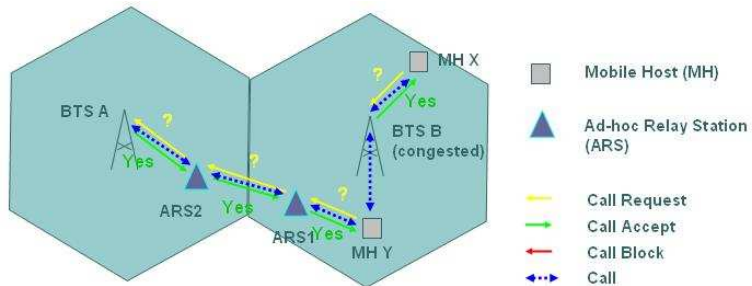


Figure 4.2: Case one of Secondary Relaying

- Or a call between two mobile hosts, one located in the congested cell and the other located either in the same cell or a neighboring cell, is relayed through ARS's allowing other mobile hosts to acquire channels.
- **Cascaded Relaying:** This type of relaying uses both previously described relaying techniques and spans multiple cells. The idea is that

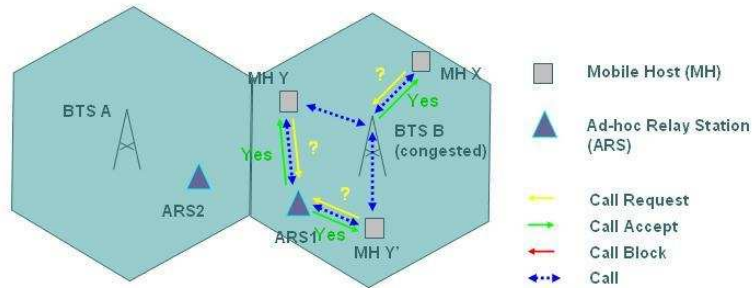


Figure 4.3: Case two of Secondary Relaying

if neither primary or secondary relaying is possible³, a neighboring cell will free up a channel through primary relaying which will then enable secondary relaying in the originating cell by using the newly freed up channel. Figure 4.4 illustrates this case, MH Z frees up a channel in cell C via primary relaying which makes it possible to free a channel in cell B for MH X using secondary relaying.

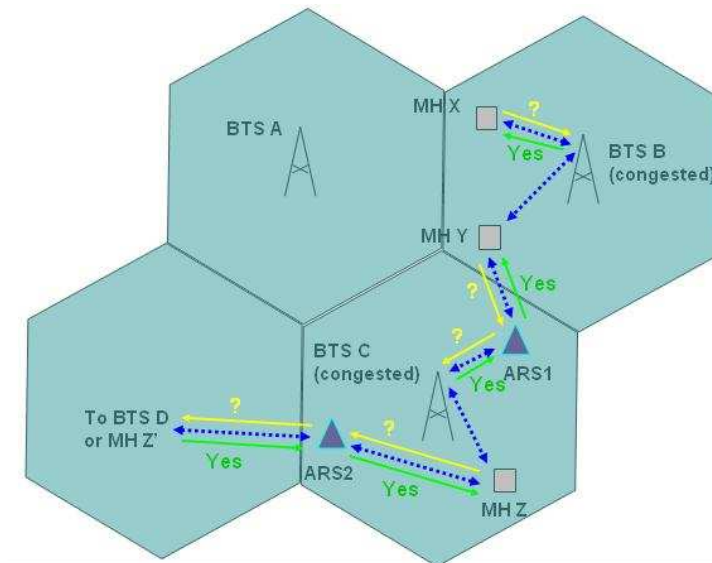


Figure 4.4: Cascaded Relaying

In the cases presented above, the reason for using an ARS is congestion in a cell. However these mechanisms can also be applied in situations mentioned earlier, where dead spots are encountered but where the host is still in the coverage area of an ARS. The main difference between the iCAR

³neighboring cells could also be congested

system and the one presented in this work is that in this work, the stations used to relay communications have the same mobility properties as regular mobile hosts.

4.1.2 UCAN

The Unified Cellular and Ad-Hoc Network Architecture is a system designed for interworking between 802.11 and third-generation wireless data networks.

In this system, a UE realizing that it currently has a low signal-to-noise ratio, and thus a low data rate, can turn to other UEs having a better SNR⁴ to forward its traffic.

In UCAN, the protocol which determines which UE has the best SNR ratio in order to forward traffic using the Ad-Hoc network is called "Proxy Discovery". Two different protocols are available: Greedy (which is proactive) and On-Demand (which is reactive). As with AODV for example, a route repair protocol is also implemented.

Figure 4.5 illustrates the use of UCAN.

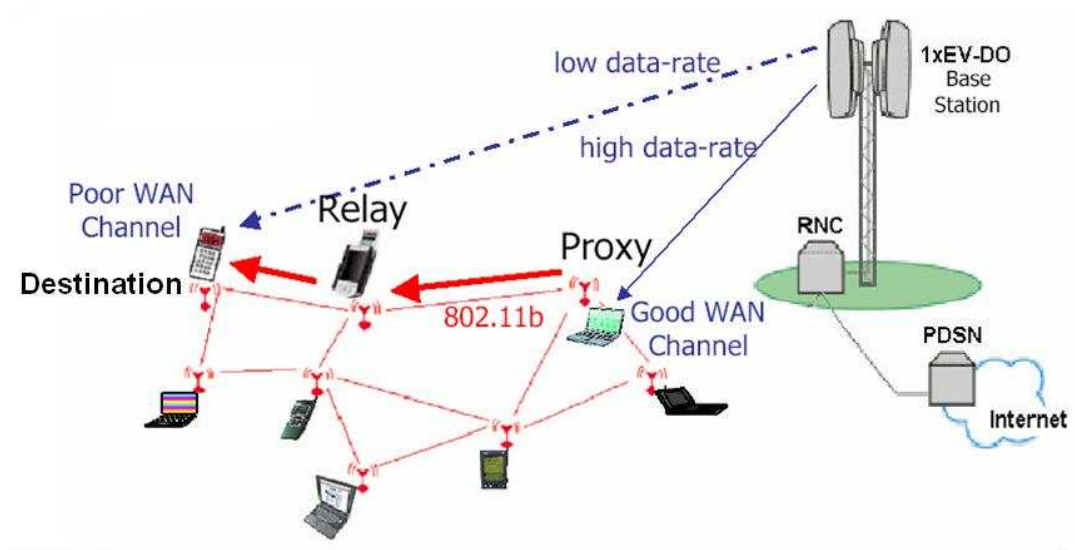


Figure 4.5: Example of UCAN use

⁴Signal-to-Noise Ratio

4.1.3 Conclusion

Much work has been done in the area of interconnecting Cellular and Ad-Hoc networks. Aside from the couple of solutions presented in this chapter, many other systems have been developed, especially for disaster crisis management. Here are some references to other interconnection systems: [23][25][24]

4.2 Motivations of this study

This study focuses on providing cellular network access to mobile hosts, in an indoor environment, which are out of range of any base station. In other words, the object of this work is to extend the coverage of a cellular network using Ad-Hoc networking.

4.2.1 Propagation Issues⁵

As we have seen in the chapter concerning cellular networks, mobile hosts in such systems can be deprived of network access if reception from a base station is impossible. This section will examine the reasons why such a situation should arise, particularly indoors.

Basic Radio Propagation

The most basic model of radio wave propagation involves the so called "free space" or "Strong Line Of Sight" radio wave propagation. In this model, radio waves originate from a source of radio energy and travel in all directions in a straight line, filling the entire spherical volume of space with radio energy that varies in strength with respect to the power of the distance⁶ (or 20 dB per decade increase in range). This means that if you double the distance over which you transmit, the received power will be reduced by a factor of 4. As an example, figure 4.6 shows the amount of power received when one Watt of power is transmitted over the air.

⁵This section was inspired by [14] and [15]

⁶Theoretical formula: $\frac{P_r}{P_e} = \left(\frac{\lambda}{4\pi d}\right)^n$ with $n \geq 2$

Transmission Range (d)	Average Power Received
1 meter	0.00002 Watts
5 meters	0.0000013 Watts
10 meters	0.0000004 Watts
1000 meters	0.000000000158 Watts

Figure 4.6: Average power received when transmitting one Watt of power using the free space model

To provide a comparison, when transmitting 1 Watt of power over a fiber optic cable that is 1000 meters long, on average, 0.933 Watts of power is received.

Real world radio propagation rarely follows this simple model. The three basic mechanisms of radio propagation are attributed to reflection, diffraction and scattering. All three of these phenomena cause radio signal distortions which in turn result in signal fades, as well as additional signal propagation losses. Outdoors, movements over very small distances by mobile hosts give rise to signal strength fluctuations because the composite signal received is made up of a number of components from the various sources of reflections (called "multipath signals[16]") from different directions as well as scattered and/or diffracted signal components. These signal strength variations amount to as much as 30 to 40 dB in frequency ranges useful for mobile communications and account for some of the difficulty presented to the designer of reliable radio communications systems. The basic signal attenuation with range noticed in the real world gives rise to what are termed "large scale" effects, while the signal strength fluctuations with motion are termed "small scale" effects.

Multipath Signals

In order for the free space model to be applicable, the antennas of communicating devices need to see each other. In urban areas and especially in indoor environments such a situation is highly improbable.

In the real world, multipath occurs when there is more than one path available for radio signal propagation. The phenomena of reflection, diffraction and scattering all add radio propagation paths to the initial direct "line of sight" path between the transmitter and receiver.

Here is the description of these three phenomena by Theodore S. Rappaport[17]:

- Reflection occurs when a propagating electromagnetic wave impinges upon an object which has very large dimensions when compared to the wavelength of the propagating wave. Reflections occur from the surface of the earth and from buildings and walls.

- Diffraction occurs when the radio path between the transmitter and receiver is obstructed by a surface that has sharp irregularities (edges). The secondary waves resulting from the obstructing surface are present throughout the space and even behind the obstacle, giving rise to a bending of waves around the obstacle, even when a line-of-sight path does not exist between transmitter and receiver. At high frequencies, diffraction, like reflection, depends on the geometry of the object, as well as the amplitude, phase, and polarization of the incident wave at the point of diffraction.
- Scattering occurs when the medium through which the wave travels consists of objects with dimensions that are small compared to the wavelength, and where the number of obstacles per unit volume is large. Scattered waves are produced by rough surfaces, small objects, or by other irregularities in the channel. In practice, foliage, street signs, and lamp posts induce scattering in a mobile communications system.

Multiple signal propagation paths are caused by any of the above phenomenon. The actual received signal level is the vector sum of all the signals incident from any direction or angle of arrival. Some signals will aid the direct path, while other signals will subtract (or tend to vector cancel). The total composite phenomenon is thus called multipath. Two kinds of multipath exist: specular multipath arising from discrete, coherent reflections from smooth metal surfaces; and diffuse multipath arising from diffuse scatterers and sources of diffraction (the visible glint of sunlight off a choppy sea is an example of diffuse multipath).

Both forms of multipath are bad for radio communications. Diffuse multipath provides a sort of background "noise" level of interference, while specular multipath can actually cause complete signal outages and radio "dead spots" within a building. This problem is especially difficult in underground passageways, tunnels, stairwells and small enclosed rooms.

Figure 4.7 presents an example of multipath signal.

- Path 1: Is a line-of-sight path. The amount of power we receive at the mobile host depends only on the distance between the two devices.
- Path 2: Is a path that is the result of a reflection off an object. Here we will lose power over the length of the path from the base station to the object, and over the length of the path from the object to the mobile host. It is very important to note that, in this case, we are not propagating over the distance between the base station and the mobile unit but over the total length of Path 2. In addition, the object will also absorb some of the energy of the signal before it is reflected. The amount that it absorbs depends on the material of that object.

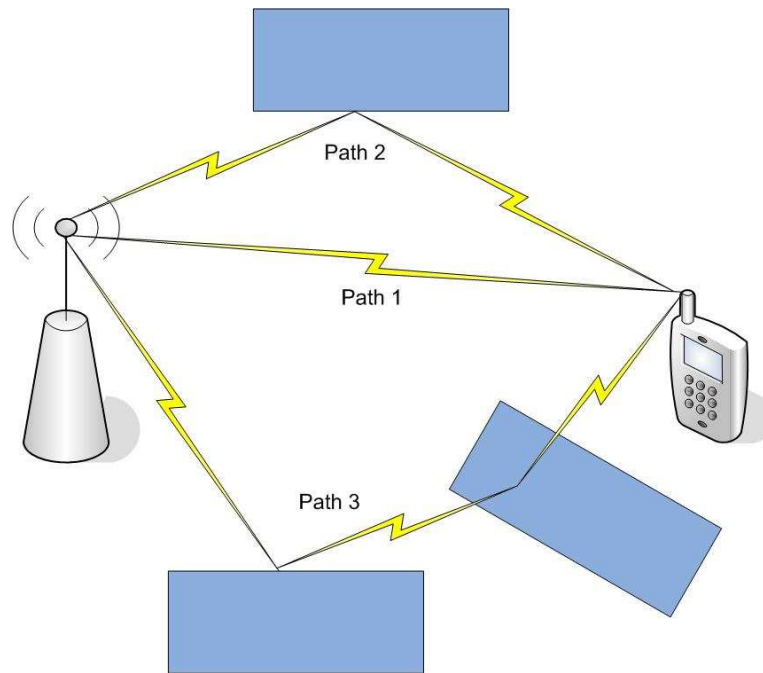


Figure 4.7: Example of multipath signal

- Path 3: Similar to Path 2, we have the signal reflecting off an object on its way from the base station to the mobile host. Notice also that it passes through another object on its way to the mobile unit. The amount of power that is reduced as it passes through the object depends again on the material of that object.

In this example the mobile will receive so much power from Path 1, since it did not reflect or pass through any object, and because it was the shortest of the three paths, that we can ignore Paths 2 and 3. However, in many situations, we do not have direct line-of-sight between devices and communication between them can only be made possible through indirect paths like 2 and 3.

Indoor Radio Propagation

Indoor radio propagation is characterized with the same effects as outdoor propagation: reflection, diffraction, and scattering. But the influence of each parameter is much greater for indoor propagation. The behavior of radio signals in buildings depends on the working frequency, the building layout, the construction materials used and the building type. Walls and obstacles such as desks and partitions, made of different materials obstruct the signal

differently. Therefore, it is very difficult to design an "RF⁷ friendly" building that is free from multipath reflections, diffraction around sharp corners or scattering from wall, ceiling, or floor surfaces (needless to say that perfect operation is nearly impossible in existing buildings). The best case scenario for an "RF friendly" building would be an all wooden or all fiberglass structure but even then such a building will still have reflections, multipath and other radio propagation disturbances which will prove to be less than ideal.

Inside smooth walled metal buildings, radio wave propagation can be so bad that "dead spots" can appear, where the signal is virtually non-existent. The reason for this is because of almost perfect, lossless reflections from smooth metal walls, ceilings or fixtures that interfere with the direct radiated signals. The dead spots exist in three dimensional space within the building, and movements of only a few centimeters can alter reception from no signal to full signal.

Radio wave propagation obstacles are divided into two categories: hard partitions if they are part of the physical/structural components of a building and soft partitions which are formed by the office furniture and fixed or movable/portable structures that do not extend to a buildings ceiling. Radio signals effectively penetrate both kinds of partitions in ways that are very hard to predict.

An obstacle with a measured loss of 20dB or more from its materials is a significant loss if you consider the free space model where 20db is the additional loss reported per decade increase in range. The equivalent to a transparent object in radio wave propagation would consist of a material with a three to six dB loss.

In order to predict the signal level inside a building, different parameters have to be considered⁸:

- Building penetration caused by the signal entering the building. It has two main factors: wall penetration and window area. It is found that the wall penetration for 900 MHz signals can be in the range 15 - 27 dB, depending on building construction. The building penetration in the window area is approximately 6 dB.
- Floor attenuation depends on the number of floors between the transmitter and receiver, the type of the material, and the working frequency. It is found that for the carrier frequency of 900 MHz, the attenuation in the first thirteen floors is nearly the same, about 2.7 dB/floor. Above thirteen floors the attenuation is approximately 7 dB/octave.

⁷Radio Frequency

⁸see [17] for a detailed description on signal attenuation inside buildings

For both factors, we can extrapolate this data for use at around 2GHz, which is used for UMTS, if we add a few dB (perhaps 5 to 6 dB) to account for the higher frequency.

Propagation Models

Two different propagation models will be used in this work:

1. The Freespace Model[18]:

The free space propagation model assumes the ideal propagation condition that there is only one clear line-of-sight path between the transmitter and receiver. H. T. Friis presented the following equation to calculate the received signal power in free space at distance d from the transmitter[20].

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} \quad (4.1)$$

where P_t is the transmitted signal power. G_t and G_r are the antenna gains of the transmitter and the receiver respectively. $L(L \geq 1)$ is the system loss, and λ is the wavelength.

The free space model basically represents the communication range as a circle around the transmitter. If a receiver is within the circle, it receives all packets. Otherwise, it loses all packets.

2. The Shadowing Model[19]:

The shadowing model consists of two parts. The first one is known as the path loss model, which also predicts the mean received power at distance d , denoted by $\overline{P_r(d)}$. It uses a close-in distance d_0 as a reference. $\overline{P_r(d)}$ is computed relative to $P_r(d_0)$ as follows.

$$\frac{P_r(d_0)}{\overline{P_r(d)}} = \left(\frac{d}{d_0}\right)^\beta \quad (4.2)$$

β is called the path loss exponent, and is usually empirically determined by field measurement. From Eqn. (4.1) we know that $\beta = 2$ for free space propagation. Larger values correspond to more obstructions and hence faster decrease in average received power as distance becomes larger. $P_r(d_0)$ can be computed from Eqn. (4.1).

The path loss is usually measured in dB. So from Eqn. (4.2) we have

$$\left[\frac{P_r(d)}{P_r(d_0)} \right]_{dB} = -10\beta \log \left(\frac{d}{d_0} \right) \quad (4.3)$$

The second part of the shadowing model reflects the variation of the received power at certain distance. It is a log-normal random variable, that is, it is of Gaussian distribution if measured in dB. The overall shadowing model is represented by

$$\left[\frac{P_r(d)}{P_r(d_0)} \right]_{dB} = -10\beta \log \left(\frac{d}{d_0} \right) + X_{dB} \quad (4.4)$$

where X_{dB} is a Gaussian random variable with zero mean and standard deviation σ_{dB} . σ_{dB} is called the shadowing deviation, and is also obtained by measurement. Eqn. (4.4) is also known as a log-normal shadowing model.

The shadowing model extends the ideal circle model to a richer statistic model: nodes can only probabilistically communicate when near the edge of the communication range.

Conclusion

As we have seen indoor radio propagation is extremely variable and tremendously tricky to predict (sometimes even impossible). Using a cellular system such as UMTS it is inevitable to encounter some dead spots (basements or parking lots are common examples) inside buildings where no reception is possible. For mobile hosts which find themselves in such situations, Ad-Hoc networking can be of valuable assistance. Instead of having mobile units deprived of any connectivity, using a Ad-Hoc approach these hosts can find a route, made up of other mobile hosts, ending with a host connected to the cellular network. This would then allow the node to enjoy the services of the cellular network even though no base station signal is available. This is called coverage extension.

Figure 4.8 illustrates this solution.

In this figure:

- Mobile host 1 has direct access to the cellular network through a base station.
- Mobile hosts 2 and 3 on the other hand have to go through host 1 to have access to the cellular network.

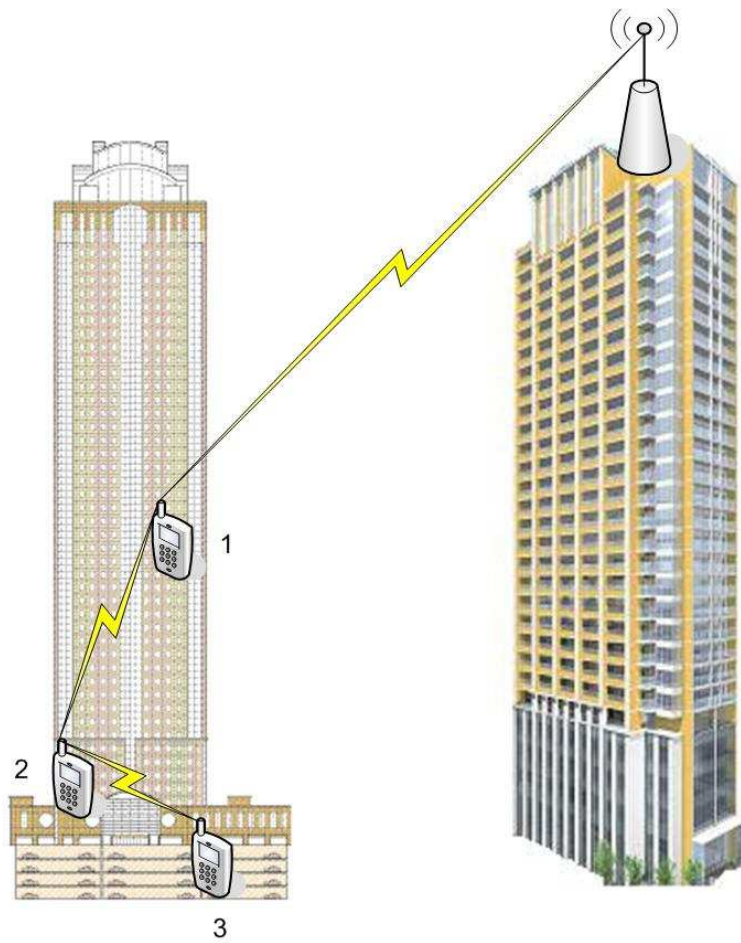


Figure 4.8: Cellular network assisted by Ad-Hoc networking

Part II

Personal Contribution

Introduction

This part will describe a new solution for extending a cellular network (i.e. a 3G network in this case) using Ad-Hoc networks and suitable for an indoor environment.

To start with, a more detailed description of AODV (see chapter 3) will be presented along with the ABR (Associativity Based Routing) protocol. Then, the hypotheses and assumptions made for this study will be presented. After that, the modifications and additions to AODV named GWAODV⁹, will be detailed and finally NS-2, the network simulator used to evaluate this new protocol along with the results it generated will be explained.

⁹GWAODV: Gateway Ad-Hoc On-demand Distance Vector

Chapter 5

AODV: Ad-hoc On-demand Distance Vector

As mentioned earlier, AODV[1] is a reactive routing protocol which means that an unknown route is discovered only when needed. However, unlike DSR, AODV makes use of beacon messages so that nodes can maintain a list of immediate neighbours. These messages are sent with a time-to-live of one to prevent distribution past immediate neighbours.

Route repair mechanisms allow routes to be reconstructed if an intermediate node goes down. This feature makes it possible for AODV to adapt to a changing network topology.

In order to avoid routing loops and count-to-infinity problems associated with classical distance vector algorithms, AODV uses sequence numbers which are included in every route information message.

5.1 Message Types

There are three different messages in AODV:

- RREQ: Route Request which is sent when a route for an unknown destination is required
- RREP: Route Reply which is sent as a reply to a RREQ either by the destination itself or an intermediate node which has an active route to a destination.
- RERR: Route Error which is sent when a node realizes that a neighbour is no longer reachable.

An additional beacon or "Hello" message is also used to notify immediate neighbours of the node's presence, these messages are periodically sent to a node's neighbours.

5.2 AODV Operation

This is a simplified description of the operation of AODV. For a complete understanding of this protocol, please see [1].

5.2.1 Sequence Numbers

A sequence number is maintained in every node; it is incremented before sending out either a RREQ or a RREP. When a route is entered in a node's routing table, it also records the destination's sequence number contained in the RREQ message. This destination sequence number has to be the latest sequence number available to prevent loops if the information is stale. A node changes the destination sequence number of a routing table entry when:

- it is itself the destination and it offers a new route to itself
- it receives a message with new information about the sequence number of the destination
- the path towards the destination expires or breaks

5.2.2 Routing Table

The routing table of a node contains the following information for each route:

- Destination IP address
- Destination Sequence Number: The destination sequence number associated with the route.
- Next Hop: The IP address of either the destination or an intermediate node along the route to the destination.
- Hop Count: The number of hops from the source to the destination.
- Flags: State of the route; valid (currently being used), invalid (not used at the moment but still in cache for information purposes) or being repaired.
- Lifetime: For an active route this field denotes the time at which the entry becomes invalid (expiry time). For invalid routes it represents the deletion time.

An entry of the routing table is created whenever a node receives a message for an unknown destination.

An entry is updated when:

- a message with a higher destination sequence number is received

- a message with the same destination sequence number is received but with a smaller hop count
- the destination sequence number is null (i.e. unknown) and a message containing a sequence number for that destination is received
- a message with the same destination sequence number is received and the route is invalid (update of deletion time)

5.2.3 Route Discovery

Creating a RREQ

A node initiates a route discovery phase by generating a RREQ. This phase is needed when a node wishes to communicate with a previously unknown destination or a destination for which its routing table entry has become invalid. As an example, in figure 5.1, node 1 wants to exchange information with node 4 but it doesn't have a route to that destination. Node 1 will

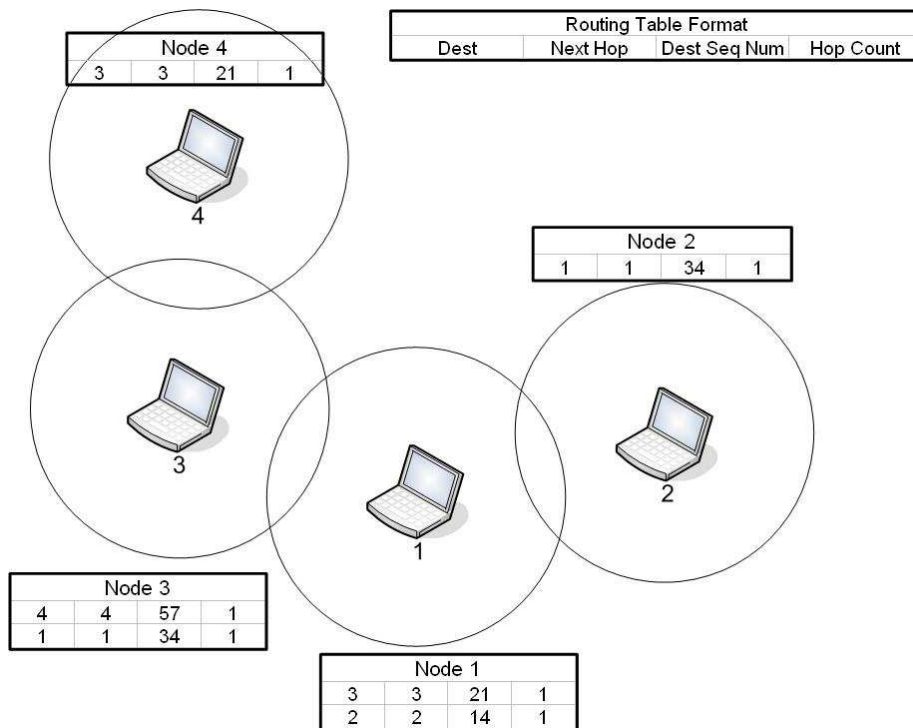


Figure 5.1: Example of the need of a route discovery phase

then construct a RREQ message containing:

- a RREQ ID which allows nodes to discard requests they have already seen, it's incremented by the source node before each new RREQ

- Destination IP address and sequence number
- Originator IP address and sequence number which is incremented before insertion
- a Hop Count of 0
- an empty list of IP address-sequence number pairs (path list) which will be updated by nodes receiving this RREQ so that the RREP message will know the route back to the originator

Once this RREQ is generated, the node saves this message in a buffer to avoid reprocessing or re-forwarding this RREQ if it receives it from a neighbour and broadcasts the message. Figure 5.2 illustrates this process.

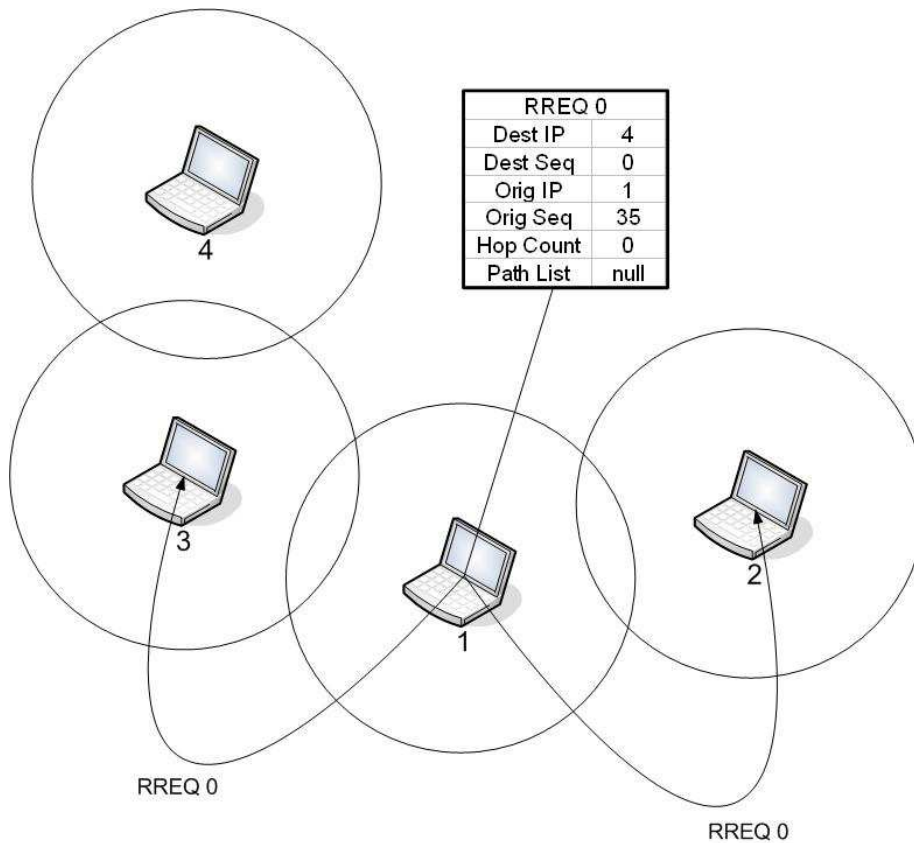


Figure 5.2: RREQ creation and distribution

Receiving a RREQ

Node 2 and 3 will receive this RREQ. Both nodes will increase the hop count field by one. As a general rule, each node receiving this RREQ may create

or update a route to any destination listed in the path list included in the RREQ. Since node 2 and 3 are neighbours of node 1, this list is empty and no modifications are made to their routing table.

Next, each node checks if they have already received a RREQ from the same source with the same RREQ ID. If such a RREQ has been received, the node discards the newly received RREQ.

Otherwise, two options are available depending on the presence or not of a route to the destination in a node's routing table. Node 2 and 3 illustrate these two possibilities.

Forwarding a RREQ

Node 2 has no routing table entry for node 4. If the time-to-live field of the IP header is greater than one then node 2 will rebroadcast this RREQ. Prior to this node 2 has to decrease the time-to-live field of the IP header by one and append its own IP address and sequence number to the path list of the RREQ.

Creating a RREP

Node 3 on the other hand is a neighbour of node 4 and its routing table contains an entry for that destination. Therefore node 3 is able to respond to the RREQ through a RREP which contains:

- the Destination Sequence Number found in its routing table
- the Destination IP address, Originator IP address and Originator Sequence number extracted from the RREQ.
- the hop count field of the routing table is copied in the RREP
- the path list of the RREQ is included in the RREP in the same order as in the RREQ

The RREP can now be unicast to next hop towards the originator, indicated by the last entry in the path list.

Figure 5.3 demonstrates this process. The RREQ forwarded by node 2 and which reaches node 1 is not represented on the figure for clarity.

Receiving a RREP

When a node receives a RREP message, it increments the Hop Count field. Then it may update or create an entry for the Destination IP (i.e. node 4 in this case) of the RREP and any other node in the path list. If the node receiving the RREP is not the destination, it has to forward the message to the previous node in the path list (the path list is read in reverse order).

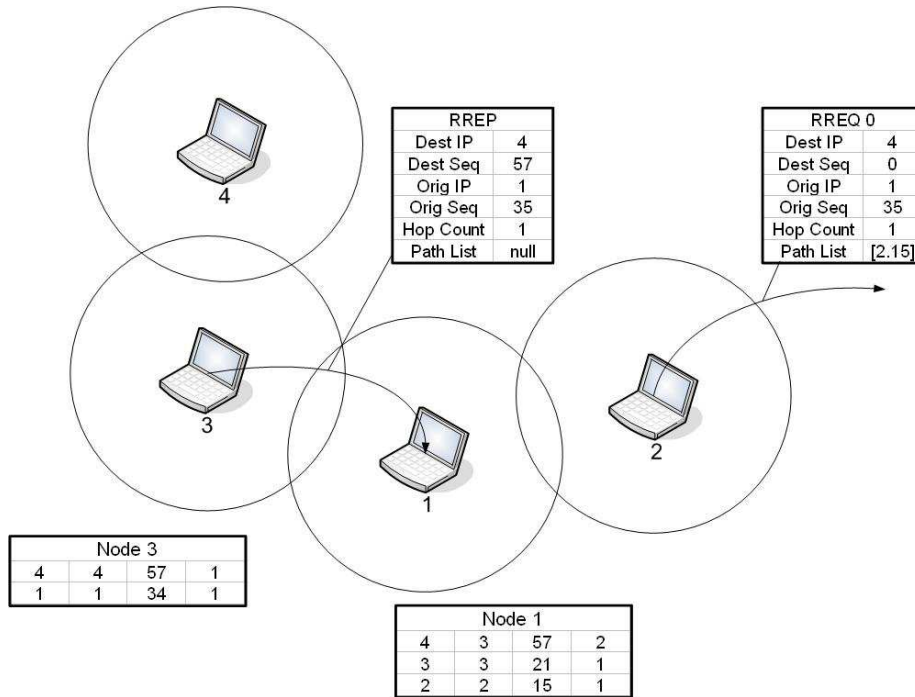


Figure 5.3: RREQ forwarding and RREP creation

In this example, node 1 increments the Hop Count and updates its routing table. It now has an active route to node 4 (See figure 5.3 for node 1's routing table).

5.2.4 Route Repair and RERR

RERR

A node generates a RERR message in three situations:

1. The node detects a link break for the next hop of an active route in its routing table while transmitting data (and route repair, if attempted, was unsuccessful).
2. The node receives a data packet for an invalid route or an unknown destination.
3. The node receives a RERR from a neighbor for one or more active routes.

This RERR message will contain a list of unreachable destinations according to the situation which arises.

For the first case, the node creates a list of unreachable destinations consisting of the unreachable neighbor and any additional destination present

in the node's routing table and using the unreachable neighbor as the next hop.

In case 2, there is only one unreachable destination, which is the destination of the data packet that cannot be delivered.

In the last case, the list should consist of those destinations in the RERR for which there exists an entry in the node's routing table that has the transmitter of the received RERR as the next hop.

In all cases, a node will invalidate any route where the destination is present in the unreachable destination list. The node will then broadcast this message to all its neighbours.

Local Repair

When a link break in an active route occurs, the node upstream of the break can choose to repair the link locally if the destination is no farther than a certain amount of hops. To repair the link break, the node increments the sequence number for the destination and begins a route discovery phase but on a smaller scale than the original RREQ¹.

If at the end of the discovery period, the repairing node has not received a RREP (or other control message creating or updating the route) it transmits a RERR message for that destination as seen in the previous section.

On the other hand, if the node receives one or more RREPs (or other control message creating or updating the route to the desired destination) during the discovery period, it can update its route table entry for that destination. The node can now use this active route to forward data packets.

Local repair of link breaks in routes enables nodes to maintain communication while the network topology changes. However this process sometimes results in an increased path length to a destination.

¹the time-to-live field is set for a smaller distribution radius

Chapter 6

ABR: Associativity-Based Routing

ABR is a reactive and beacon-based routing protocol. It was developed by C.K. Toh at Cambridge University in 1996. This protocol selects routes based on the temporal stability of the links between the nodes. The fundamental objective is to find longer-lived routes.

Each node generates periodic beacons (hello messages) to signify its existence to its neighbors. These beacons are used to update the associativity table of each node. This table contains the associativity level between a node and its neighbours. With the temporal stability and the associativity table the nodes are able to classify each neighbor link as stable or unstable.

Stability is determined using "associativity ticks". Association in ABR is based on a few metrics such as link delay, signal strength, power life, route relaying load, period of presence or spatial and temporal characteristics. Routes are only chosen when they have a high degree of associativity which means a high level of associativity ticks.

6.1 ABR Operation

The ABR protocol consists of three phases; route discovery, route reconstruction and route deletion.

6.1.1 Route Discovery

If a node has in his Route Cache a route to the desired destination then this route is immediately used. If not, the Route Discovery protocol is started.

First the network is flooded with RouteRequest messages originating from the source. These messages are only forwarded once by each intermediate node. When receiving a RouteRequest message, intermediate nodes append their address and associativity ticks to the packet.

Once the RouteRequest message reaches the destination node, this node will wait a certain period of time during which it may receive other RouteRequest messages from the same source but delivered along different routes. It will then select the best route by examining the associativity ticks along each path. If multiple routes have the same overall degree of stability, the route with the minimum number of hops will be selected.

After the route has been chosen, the destination sends a Reply packet back to the source along the same path.

Figure 6.1 illustrates this process.

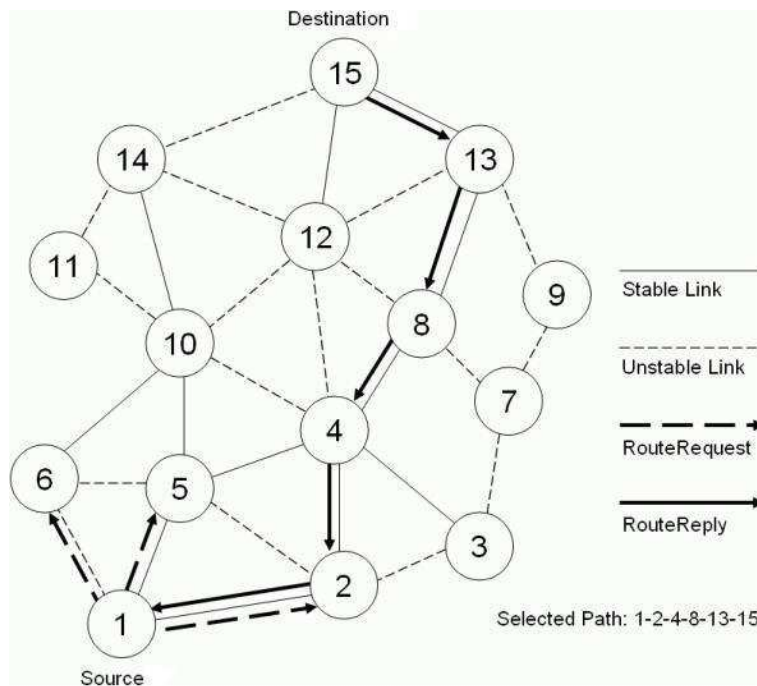


Figure 6.1: ABR Route Discovery

Three routes are possible from 1 to 15:

- 1-5-10-14-15
- 1-5-4-12-15
- 1-2-4-8-13-15

ABR selects the last route because of the highest percentage of stable links on this path.

6.1.2 Route Repair

As with Local Repair in AODV, if a link is broken, a localized reconstruction phase is started. However this procedure is more complex than the one

implemented in AODV. The different operations which are executed depend on which node causes the link break. In ABR the route reconstruction phase takes into account movements by the source, destination, intermediate and concurrent nodes. Each node category determines a different set of actions. Since these concepts are not used in GWAODV they will not be described here but more information is available in [2].

6.1.3 Route Delete

If a discovered route is no longer needed, the source node initiates a RouteDelete broadcast. All nodes along the route, delete the route entry from their routing table. The RouteDelete message is fully broadcasted, because the source is not aware of any changes in the path which may have been caused by a RouteRepair.

Chapter 7

Hypotheses and Assumptions

As seen earlier ¹ this work presents a solution allowing out of range UEs to maintain an access to the cellular network using a series of other mobiles nodes. The last mobile station before a NodeB will be referred to as a gateway UE.

7.0.4 User Equipment (UE)

The mobile units used in the simulations do not differ from each other, they all possess two interfaces: an UMTS and an 802.11. Since these are mobile units, any UE can at one point act as a gateway as long as the NodeB allows it. A mobile unit functioning in Ad-Hoc mode and entering the vicinity of a NodeB switches to its UMTS interface.

Also, a gateway UE cannot refuse traffic from another UE. We consider that they can forward every packet.

7.0.5 UMTS model

The UMTS model used in this work is highly simplified and consists of one NodeB connected to a wired network with an unlimited number of UEs. Entities such as the SGSN, GGSN, HLR, VLR etc. have not been modeled to keep this study simple and because the focus of this work is on the interactions between mobile stations. The NodeB has no knowledge of the Ad-Hoc network topology hidden behind the gateway nodes it can only allow or prevent a UE from acting as a gateway for other nodes.

7.0.6 Traffic Direction

The simulations carried out in this work rely on data transmitted from the UEs to the cellular network, the traffic in the opposite direction has not been considered. An example solution for traffic originating from the

¹See figure 4.8 for a usage scenario

cellular network and intended for a UE out of reach of a NodeB could be implemented using encapsulation.

The source encapsulates packets, adding a header containing the destination address of a gateway UE located in the same Ad-Hoc network as the destination. Once the gateway UE receives a packet it decapsulates it and forwards it to the destination using an Ad-Hoc routing protocol. The question of how this source knows to which gateway UE it has to send its packets can be answered simply by assuming that this node will have previously received a packet from the destination through the appropriate gateway.

Since the solution presented here is based on an existing Ad-Hoc routing protocol, it is able to support "inter-Ad-Hoc" network communications. However to keep the study simple, this kind of traffic has not been considered in simulations.

7.0.7 Security and Authentication

These issues are not addressed in this study but nevertheless they mustn't be overlooked. As with all decentralized networks where registration is not required, authentication and security are complex matters, out of the scope of this work. More information is available in [26].

7.0.8 Charging and Billing

Charging and billing are essential for any commercial use. Possible solutions in this context include:

- Having the gateway take care of these issues; the cellular network provider would charge and bill the gateway as if every communication belonged to its own traffic.
- Having the cellular network reward gateways for forwarding traffic with price reductions on a monthly service plan or other benefits [27].
- Allowing the NodeB to isolate each connection forwarded by a gateway in order to charge and bill users independently.

More information is available in [28].

Chapter 8

GWAODV: Gateway AODV

8.1 Introduction

This protocol is the object of this study. Its purpose is to extend the coverage of an existing cellular network using Ad-Hoc networking. The basic idea is for each node to maintain a table of gateway nodes which give them access to the cellular network. This protocol is based on concepts taken from both AODV and ABR.

GWAODV is a proactive routing protocol with regards to maintaining a gateway table but as opposed to a protocol like OLSR flooding is not used. Instead, standard AODV beacon messages transport extra information concerning gateways. Routing inside the Ad-Hoc network is handled by AODV mechanisms although this kind of traffic is not considered in the simulations presented later it is nonetheless supported by this protocol.

When access to the cellular network is required and no base station is within range, a node chooses an entry in its gateway table and sends its packets along this route. The gateway will then relay these messages to the cellular network. The choice of the gateway depends on several metrics and the needs of the source application.

8.2 Hello-Messages

Hello-Messages are beacon control messages sent at regular intervals between neighbours. Two kinds of Hello-Messages are used in GWAODV: standard Hello-Messages as seen in AODV and Gateway Hello-Messages which are regular Hello-Messages with added gateway information.

8.3 Neighbour Table and Associativity

Each node maintains a neighbour table containing the following fields:

- Neighbour IP Address
- Associativity Level
- Missed Hello Message Count
- Active Time
- Expire Time

As long as a node receives Hello-Messages on a regular basis from a neighbour, the associativity level between these two nodes will grow until a maximum value¹ is reached. This process is depicted in figure 8.1.

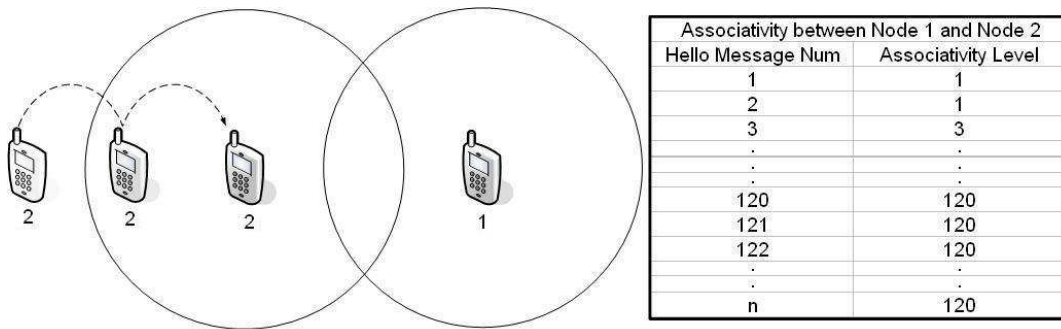


Figure 8.1: Increasing Associativity Level

However, if a node doesn't receive a Hello-Message from a neighbour within an Active Time period², the Associativity Level with that neighbour is decreased by one and the Missed Hello Message Count is incremented. If the Missed Hello Message Count reaches more than five, the Associativity Level is decreased by an extra [Missed Hello Message Count] ticks. The Hello Message Count is reset to zero whenever a Hello Message is received. Figure 8.2 displays a situation where node 2 moves away from node 1 and eventually the associativity level between these nodes drops to 0.

If no Hello-Message has been received from a neighbour within Expire Time, the corresponding entry is deleted from the neighbour table.

Gateway nodes are also required to maintain the associativity level between themselves and the cellular network base station (a NodeB in this case).

¹This value, MAX_ASSOC is set to 120 in this implementation

²Hello-Messages are lost when nodes move away from each other or when a node goes down

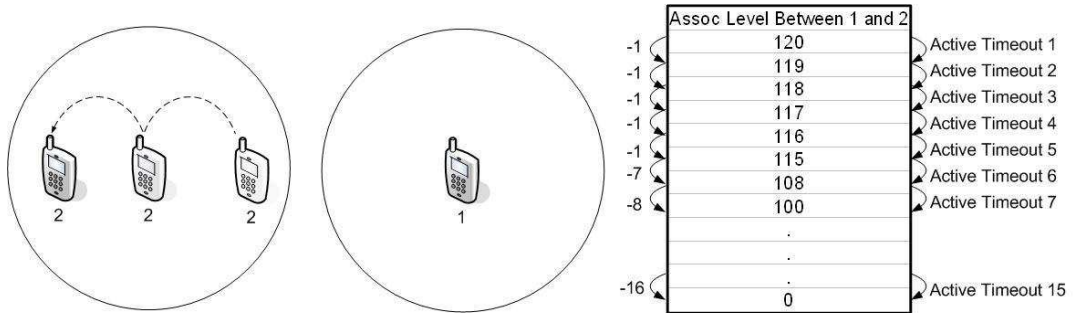


Figure 8.2: Decreasing Associativity Level

8.4 Metrics

Nodes in GWAODV rely on various metrics when selecting a gateway route. Using information contained in Gateway Hello-Messages received from neighbours, a node computes the:

- Hop Count of the entire route to a gateway
- Average Associativity Level along the route
- Standard Deviation[29] of the Associativity Level along the route
- Associativity Level between the gateway and its NodeB³

8.5 Gateway Table

Every node keeps track of a gateway table containing an entry for every gateway route learned through Gateway Hello-Messages. An entry is also inserted if the node has direct connectivity to the cellular network.

A gateway entry consists of the following fields:

- Entry ID
- Gateway IP Address
- Average Associativity Level along the route
- Standard Deviation of the Associativity Level along the route
- Associativity Level between the gateway and its NodeB
- Next Hop IP Address

³This parameter may have a different weight from other associativity levels along the route to give more or less importance to the associativity with the NodeB

- Next Hop Entry ID
- Associativity Path List
- Update Flag
- Expire Time

The Entry ID is a simple counter incremented before every insertion in the gateway table. The Next Hop Entry ID is the Entry ID representing the same gateway route in the gateway table of the next hop. Since a node can receive information about a gateway through different paths, multiple entries for a same gateway can be present in the nodes gateway table. For this reason, the Next Hop Entry ID is necessary to insure that a packet follows the intended route. It acts as a path identifier.

The Associativity Path List contains pairs of (Node-Address, Associativity-Level⁴). It is needed to compute the standard deviation along the route and to insure loop freedom as will be seen later.

The Update Flag informs a node if the gateway entry has been refreshed since sending its last Hello-Message. If Update Flag is true, the node can include the entry in a Gateway Hello-Message, if not and if no other gateway entry is fresh then a standard Hello-Message will be sent.

The Expire Time serves the same purpose as in the Neighbour Table.

8.6 Gateway Selection Algorithm

Every time a gateway route is needed, the gateway selection algorithm is invoked and returns an entry in the gateway table. The choice of the entry depends on the information gathered about the route and the context in which the call was made (i.e. choosing a route when sending a Gateway Hello-Message or when an application needs to send data packets).

8.6.1 Algorithm Operation

To begin with, every gateway route is a potential candidate. Then each metric is considered one after the other. The order is determined by the parameters passed to the algorithm. For each metric, only the best routes with an added tolerance level are kept.

For instance, if the metric is hop count and the tolerance level is two hops, the algorithm will keep only the routes with shortest hop count plus the routes which surpass the shortest hop count by a maximum of two hops. Once all the metrics have been considered the first route remaining in the list is chosen.

⁴Aossocitivity Level with next hop along the route

The algorithm is called with the following parameters. the order in which the metrics are considered and the tolerance level associated with each metric.

8.7 Creation of Gateway Hello-Messages

A Gateway Hello-Message is generated instead of a standard Hello-Message whenever a node's gateway table contains at least one fresh entry (Update Flag set to true). In this implementation, the gateway selection algorithm is called with hop count as first metric and a tolerance of two hops, then average associativity level and a ten tick tolerance and finally standard deviation with a tolerance of ten.

Once the gateway entry has been selected the message can be created with the following fields:

- Gateway IP Address
- Entry ID
- Associativity Path List
- Vector of 5 Obsolete Gateways

Along with the fields already present in a standard Hello-Message. The Update Flag is set to false and the message can be broadcasted to the node's neighbours.

The role of the obsolete gateway list is to allow downstream nodes to delete expired gateway routes faster than waiting for the expiry time.

8.8 Reception of Gateway Hello-Messages

The first thing a node does when receiving such a message is check if its own address is not already present in the Associativity Path List. If the node belongs to the Associativity Path List, the message is discarded to avoid creating routing loops. Figure 8.3 shows how this problem can arise.

The already existing sequence numbers of AODV were not used for this problem, the reason is that a neighbour could forward a gateway route with a lower sequence number but with a different path from a previously received route with the same gateway address. If sequence numbers were used this message would have been discarded even though it could have information about a more stable route for instance⁵.

⁵The message could have traveled a longer path, thus arriving later than a newer message traveling a shorter path, but the average associativity level along this path could be higher than the current gateway table entry

After checking for cycles the node will verify if it already has a gateway table entry matching: the gateway address, the path list and the next hop of the received message. If such is the case then the node will update this gateway entry. Otherwise a new entry will be created.

The node will then add the address and associativity level of the neighbour through which he received the message to take into account the last hop and compute the different metrics which will be included in the gateway table entry.

Since this is a Hello-Message, the node will update the associativity level with its neighbouring node prior to adding it to the associativity path list.

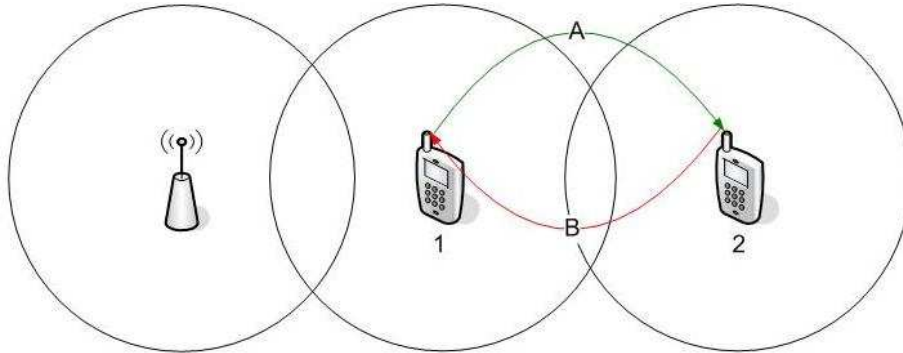


Figure 8.3: Routing loop problem

In figure 8.3 Node 1 sends a Gateway Hello-Message to Node 2 informing it that a route to a gateway is available. Node 2 receives this message and inserts a new entry in its gateway table with Node 1 as next hop.

When the next beacon message is sent by Node 2 it will include information about this newly discovered gateway. Node 1 being a neighbour will receive this message and insert a new entry in its gateway table with Node 2 as next hop. A routing loop is thus created by this new gateway table entry.

8.9 Deleting Gateway Routes

A node deletes a gateway table entry whenever it has expired, the next hop along the route expires or a list of obsolete gateways has just been received.

Each node keeps track of an obsolete gateway list which is updated at every gateway route deletion. The last five entries are sent in Hello-Messages. Once the Hello-Message has been sent, the node deletes these entries from the list.

8.10 Illustration

8.10.1 Hello-Messages

This section will illustrate the principals described above.

In figure 8.4 we can see that Node 1 has direct connectivity with the NodeB thus it broadcasts a Gateway Hello-Message. Node 2 on the other hand broadcasts a regular Hello-Message since its gateway table is empty.

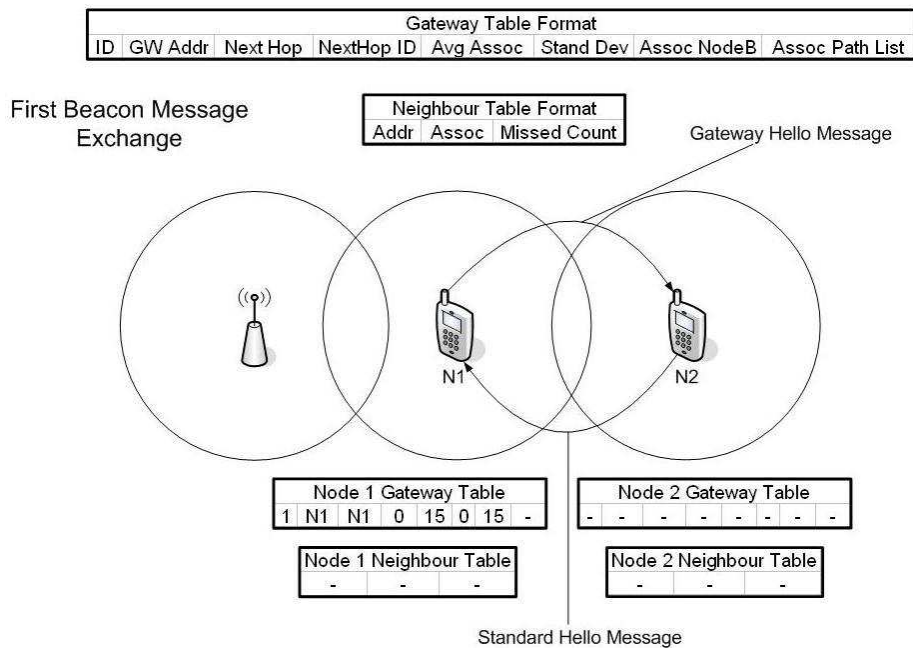


Figure 8.4: First Hello-Message Exchange

Figure 8.5 displays the state of the different tables after the first Hello-Messages have been received. The changes from figure 8.4 are marked in blue.

Node 2 has discovered a neighbour and a gateway route so both tables are updated. Since this is the first Hello-Message received from Node 1, the associativity between both nodes is set to one. The average associativity and standard deviation are computed to complete the entry.

Node 1 also updates its neighbour table since it receives Node 2's Hello-Message. This figure also suggest that the associativity between Node 1 and the NodeB has also increased.

8.10.2 Gateway Selection

Given the gateway table of figure 8.6 two examples will be presented here.

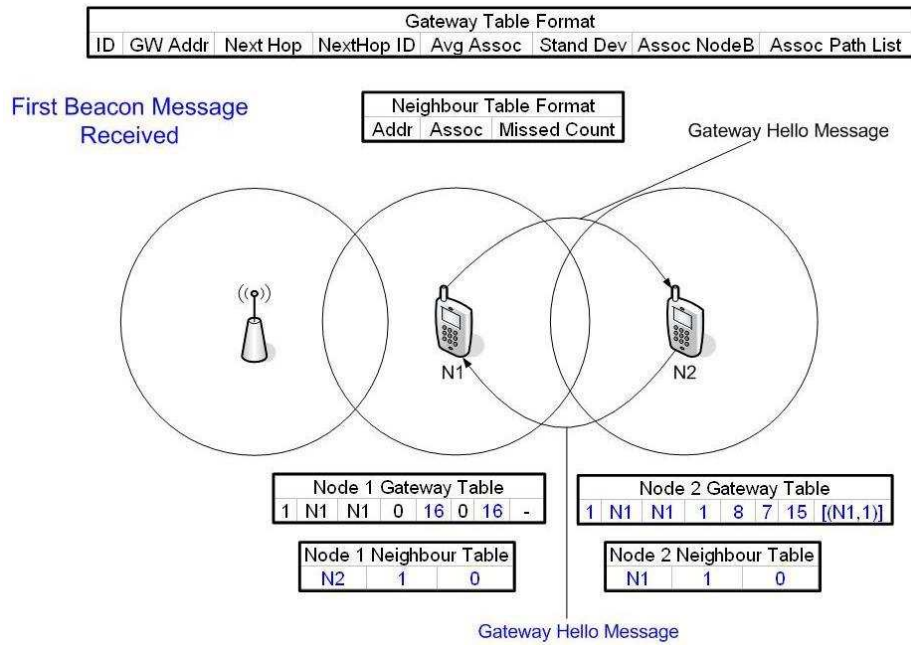


Figure 8.5: First Hello-Message Received

Node 5 Gateway Table							
ID	GW Addr	Next Hop	NextHop ID	Avg Assoc	Stand Dev	Assoc NodeB	Assoc Path List
1	N1	N3	1	19.5	5.59	27	[(N3,12),(N14,17),(N1,22)]
2	N9	N4	1	15.17	6.62	12	[(N4,9),(N7,23),(N8,6),(N19,18),(N9,23)]
3	N12	N4	2	28.8	25.05	32	[(N4,9),(N7,23),(N10,5),(N12,75)]
4	N34	N15	8	35.5	19.2	64	[(N15,42),(N23,20),(N34,16)]
5	N6	N3	2	30	13.64	45	[(N3,12),(N6,33)]
6	N12	N2	1	33.25	33.68	90	[(N2,4),(N21,26),(N12,13)]

Figure 8.6: Gateway Table

1. Node 5 is about to send a Hello-Message, its gateway table is not empty so it needs to pick a entry which will be included in a Gateway Hello-Message.

As seen earlier the algorithm is called with the following parameters: first metric is hop count, tolerance is two hops, second metric is average associativity, tolerance is 10 ticks, third metric is standard deviation and tolerance is 10.

Initially every route is a potential candidate (entries 1 to 6).

- After analyzing the first metric, the routes remaining are: 1, 3, 4, 5 and 6.
 - Next the average associativity of route is considered, the remaining routes are: 4, 5 and 6
 - Finally standard deviation will determine the chosen route: 5
2. For the second example we suppose that an application on node 5 requires a stable connection and focuses less on delay (i.e. live score of a basketball game).

The parameters for the algorithm are: first metric is average associativity, tolerance is ten ticks, second metric is standard deviation, tolerance is seven, third metric is hop count and tolerance is five.

Initially every route is a potential candidate (entries 1 to 6).

- After analyzing the first metric, the routes remaining are: 3, 4, 5 and 6.
- Next the standard deviation along a route is considered, the remaining routes are: 4 and 5
- Finally the hop count will determine the chosen route: 5

8.11 Data Packet Processing

When a data packet needs to be sent to an address outside of the Ad-Hoc network, a gateway entry is chosen following the process seen earlier and the packet is sent along that route.

In order for nodes to know that the packet they just received is destined for a gateway, a extra field was added to the packet's header containing the next hop's gateway entry ID. When a node receives such a packet it just checks the header for this new field, finds the next hop along the route in its gateway table and forwards the packet.

8.12 Gateway Route Repair

The solution presented in this study implements a simple process for gateway route repair. When a data packet reaches a node with an unknown gateway table entry ID (the route may have been deleted recently) the node will simply forward the packet along a different gateway route.

An improvement not implemented here could have been to send a message back to the source node alerting him that its gateway route has expired and that a new route is being used. At this point, the source node could choose to either keep on using the new route or use another gateway entry if the new route does not respect the requirements of the source application.

Chapter 9

NS-2: Network Simulator

9.1 Introduction

To evaluate the performance of this new protocol, multiple simulation scenarios needed to be run. The framework used for these simulations is NS-2¹ developed at UC Berkeley.

The simulator is part of the VINT project which is a DARPA-funded research project whose aim is to build a network simulator that will allow the study of scale and protocol interaction in the context of current and future network protocols. VINT is a collaborative project involving USC/ISI, Xerox PARC, LBNL, and UC Berkeley.

An important aspect of NS-2 is that its code is open source. This enables anyone to modify the simulator according to the needs of the desired simulations. This feature has for consequence that when documentation is present (some features are not documented) it is often incomplete or not up to date with the latest modifications. However, a manual documenting the essential aspects of the simulator is maintained by the VINT project.

9.2 Simulator Design

This section was inspired by a tutorial called "NS by example" by Jae Chung and Mark Claypool which can be found at: "<http://nile.wpi.edu/NS/>" and the official NS-2 manual, "The ns Manual (formerly ns Notes and Documentation)" located at: "<http://www.isi.edu/nsnam/ns/doc/index.html>"

NS-2 is an object-oriented, discrete event driven, network simulator written in C++ with an OTcl² front-end used to execute user's command scripts.

¹NS-2: The Network Simulator - ns-2, <http://www.isi.edu/nsnam/ns/>

²Tcl script language with Object-oriented extensions developed at MIT

9.2.1 Network Components

The simulator is based on a class hierarchy in C++ and a similar class hierarchy within the OTcl interpreter. The two hierarchies are closely related to each other; from the user's perspective, there is a one-to-one correspondence between a class in the interpreted hierarchy and one in the compiled hierarchy.

Users create new simulator objects through the interpreter; these objects are instantiated within the interpreter, and are closely mirrored by a corresponding object in the compiled hierarchy. Figure 9.1 illustrates this concept while figure 9.2 displays a partial view of the class hierarchy.

The one-to-one correspondence is not entirely accurate since some classes only belong to one hierarchy as can be seen in both figures, 9.1 and 9.2.

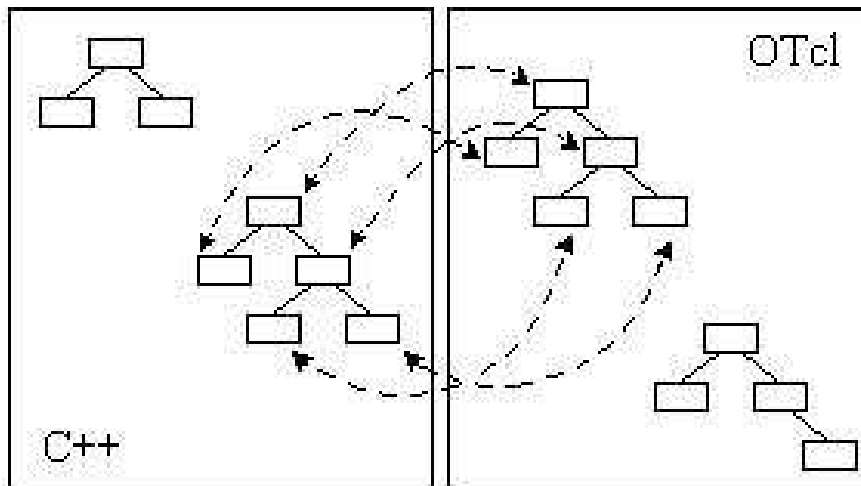


Figure 9.1: Correspondence between the C++ and OTcl hierarchies

The compiled C++ hierarchy is used to achieve efficiency in the simulations and faster execution times. This is useful for the detailed definition and operation of protocols and reduces packet and event processing time. C++ compiled objects include: a Link Layer class, a MAC 802.11 class, a Channel class and other network components.

On the other hand OTcl scripts define: a particular network topology, the specific protocols and applications to be simulated (whose behaviour is defined on the C++ hierarchy) and the form in which the simulator outputs the results. Objects in OTcl can also be used to connect multiple network components together as will be seen later, this is called "plumbing".

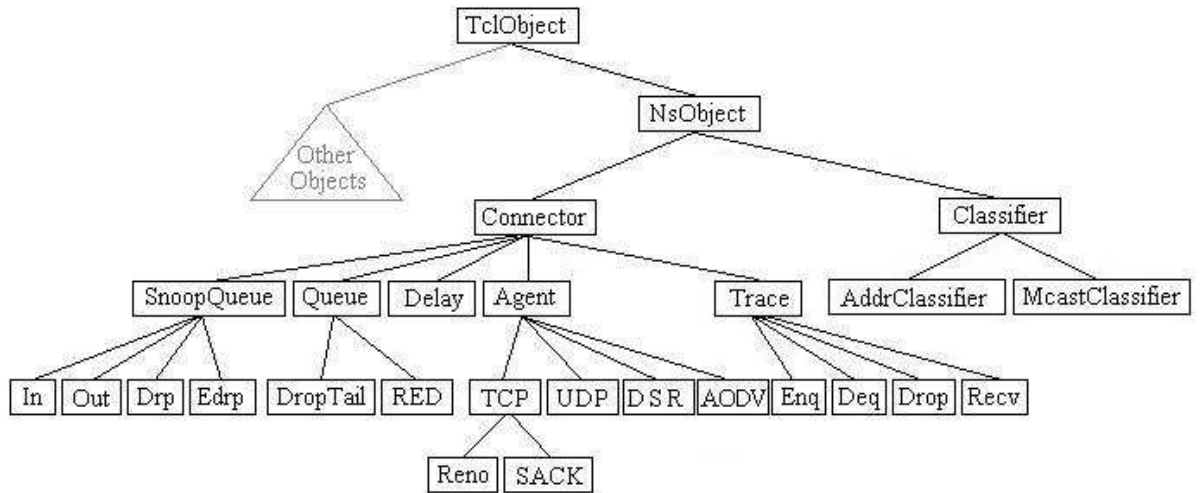


Figure 9.2: Partial class hierarchy

9.2.2 Event Scheduler

In the simulator, the advance of time depends on the timing of events which are maintained by a scheduler. An event is an object belonging to the C++ hierarchy. Each event has a unique ID, a scheduled time and a pointer to an object which will handle the event. The scheduler keeps an ordered data structure³ containing the events to be executed and fires them one by one at the appropriate time, invoking the handle method of the object pointed by the event.

The main users of an event scheduler are network components that simulate packet-handling delay or that need timers as seen in figure 9.3.

In this figure, the MAC layer object inserts an event into the scheduler. As an example, this event could be the transmission of a packet from the MAC layer to the Link layer. To account for the handling delay at the MAC layer, the event is scheduled to fire after a "Time" period. A pointer to the object handling this event, the Link layer object, is also provided by the Mac layer object.

Once the scheduler fires the event, the handle method of the Link layer object will be called and the object can start processing this event.

9.2.3 Wireless Node

A wireless node, derived from the node class is an OTcl object which connects together all the components of the network architecture described by

³By default NS-2 uses a simple linked-list

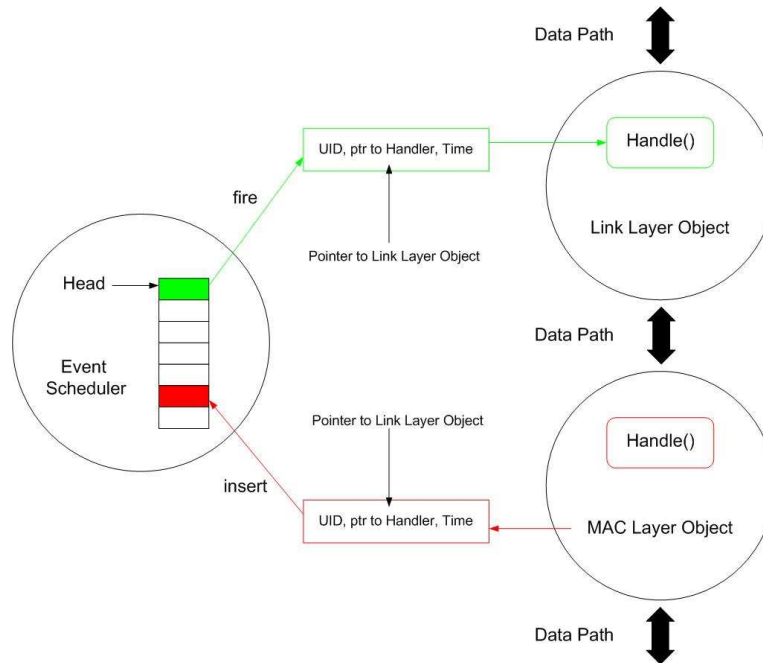


Figure 9.3: Event Scheduler

the OSI reference model minus the Session and Presentation layers. The Model used is similar to the one presented in "Computer Networks" from Andrew S. Tanenbaum[14]. Figure 9.4 represents a wireless node in NS-2⁴.

In a simulation script a user can decide the kind of components needed for each layer. A convenient interface has been developed for the first three layers, the Transport and Application layer components are added using different methods. An example configuration is given in figure 9.5.

OTcl code of the class wireless node will then create and connect each component as seen in figure 9.4.

This is an example of "plumbing" since a wireless node is a OTcl object made of C++ components.

9.2.4 Packet

A packet in NS-2 is composed of a stack of headers, and an optional data space (see Figure 9.6). A packet header format is initialized when a Simulator object is created. This defines a stack of all registered headers, such as the common header (CMN header) used by every object, the IP header, the UDP header, the AODV header and trace header, and the offset of each

⁴The Data Link Layer of the model is composed of two sub layers, the MAC (Medium Access Control) layer and the Link Layer

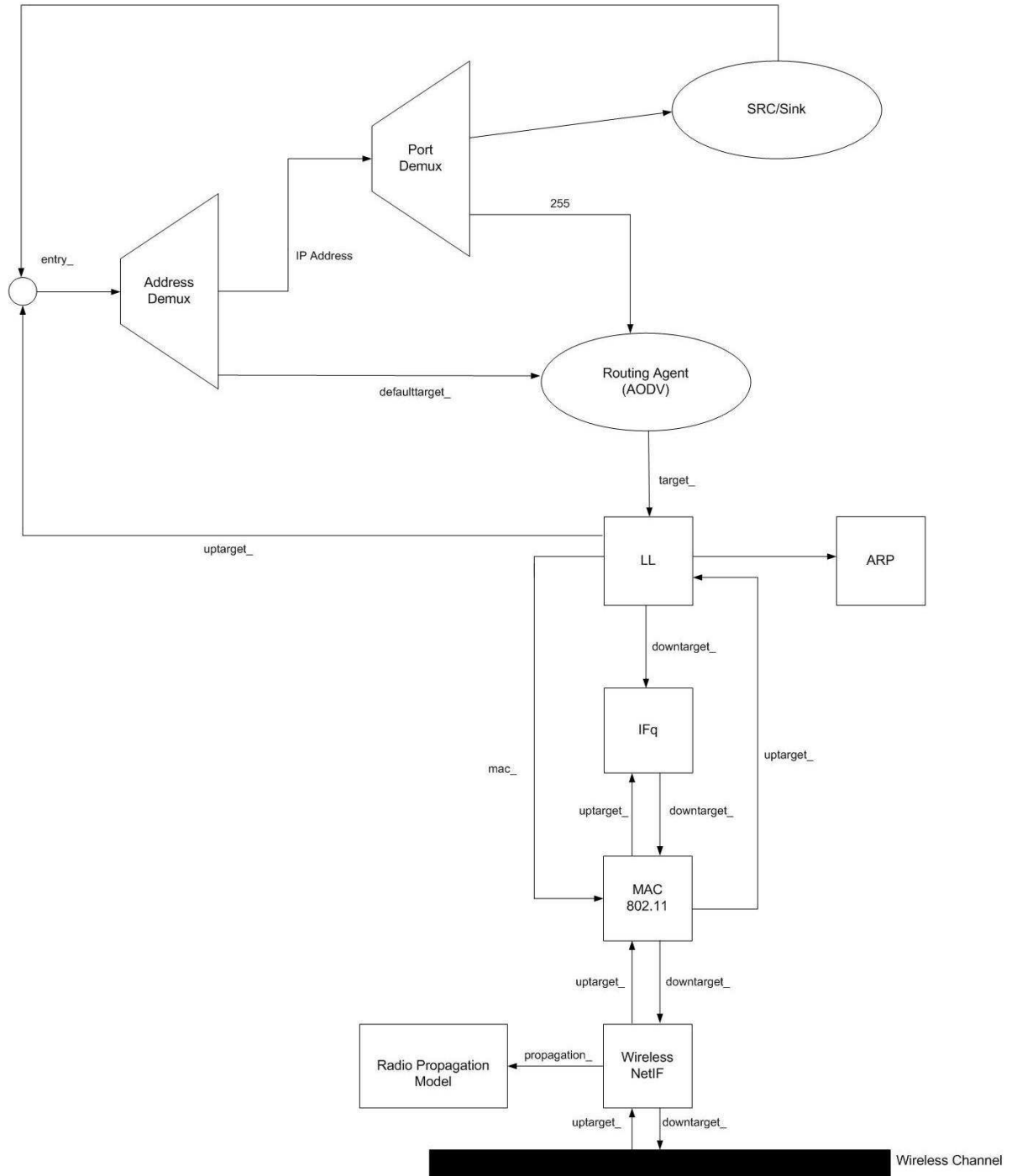


Figure 9.4: Schematic of a wireless node

Physical, Data Link and Network Layer:

```
$ns_ node-config -adhocRouting AODV \
                 -llType       LL \
                 -macType      Mac/802_11 \
                 -ifqType     Queue/DropTail/PriQueue \
                 -ifqLen      1000 \
                 -antType     Antenna/OmniAntenna \
                 -propType    Propagation/Shadowing \
                 -phyType     Phy/WirelessPhy \
                 -channelType  Channel/WirelessChannel \
                 -topoInstance $topo \
```

Where \$topo is defined as: set topo [new Topography]

Transport And Application Layer:

```
set tcp [new Agent/TCP]
set sink [new Agent/TCPSink]
$ns_ attach-agent $node_(0) $tcp
$ns_ attach-agent $node_(1) $sink
$ns_ connect $tcp $sink
set ftp [new Application/FTP]
$ftp attach-agent $tcp
```

Where TCPSink is the recipient of tcp traffic

Figure 9.5: Node Configuration

header in the stack is recorded. What this means is that whether or not a specific header is used, a stack composed of all registered headers is created when a packet is allocated, and a network object can access any header in the stack of a packet it processes using the corresponding offset value.

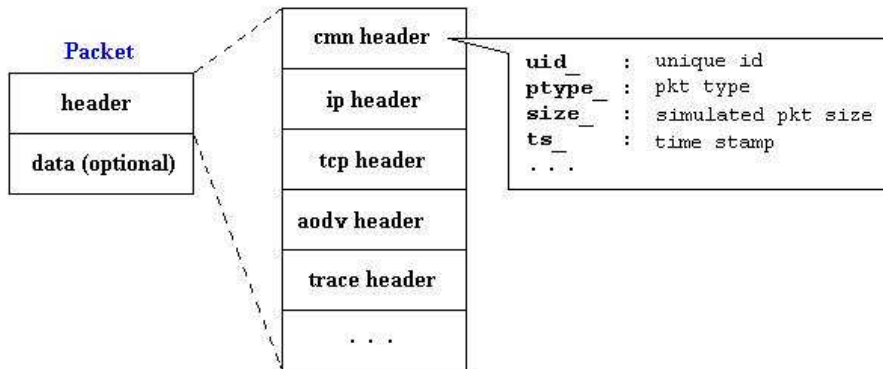


Figure 9.6: Packet Format

9.2.5 Agent

Agents represent endpoints where network layer packets are generated or consumed, they implement communication protocols at various layers. All routing protocols such as DSDV, DSR and AODV are implemented as agents but also the Transport layer protocols TCP and UDP and various applications like CBR (Constant Bit Rate) and FTP (File Transport Protocol). Agents are attached to nodes using the methods shown in figure 9.5.

Chapter 10

Simulations

10.1 Configuration

The version of the simulator used in this work is ns-2.29 which can be found at: "<http://www.isi.edu/nsnam/ns/ns-build.html>". An UMTS-FDD (FDD: Frequency Duplex Division) implementation has also been added to the simulator.

Several UMTS modules are available for NS-2, this study uses an implementation developed by the Networking Group in the INFOCOM Department at the University of Roma "La Sapienza" which can be found at: "<http://net.infocom.uniroma1.it/>".

This module was preferred to EURANE (Enhanced UMTS Radio Access Network Extensions) developed within the European Commission 5th framework project SEACORN because EURANE doesn't allow node mobility.

From <http://net.infocom.uniroma1.it/downloads/README>:

UMTS Modules for NS

These UMTS modules have been developed by Alfredo Todini and Francesco Vacirca at the INFOCOM department, University of Rome "La Sapienza", Rome, Italy.

The code began as a modification of the GPRS package by Richa Jain of the Indian Institute of Technology, Bombay, India.

The MAC and physical layers have been rewritten from scratch; the RLC module instead is based on the GPRS code, but it has been extensively modified, both to overcome its deficiencies and to make it work according to the 3GPP specifications.

Some files in the NS-2 distribution have also been modified. As in the GPRS module, it was necessary to add the NOAH (NO Ad-Hoc Routing Agent) routing algorithm.

The archive umts.tgz contains all the necessary files for ns 2.1b9a. It should be applied to the ns-2.1b9a all-in-one distribution.

10.2 Modifications to NS-2

First of all, since the UMTS module was originally developed for ns version 2.1b9a, modifications within the UMTS module were required to install this extension.

10.2.1 Double Interface Node

Every mobile terminal used in the simulations are capable of communicating with both an UMTS network and an Ad-Hoc network using the 802.11 standards. Since this kind of node didn't exist in ns-2.29 a new type of node supporting both interfaces was developed. Figure 10.1 gives a schematic view of the double interface node used in the simulations.

New fields have been added to the node configuration interface to support the addition of the 802.11 component. The added fields are: an Ad-Hoc routing agent (AODV in this case), a Link layer, an interface queue, a MAC layer, a Physical layer, a propagation model and a channel.

In the OTcl code of the simulator (ns-lib.tcl), the add-interface method of a node is called twice, the first call is made with the UMTS interface components as parameters and the second with the 802.11 interface components.

10.2.2 Modifications to AODV

AODV has been modified to incorporate the mechanisms described in the GWAODV chapter. The main modifications are in aodv(cc,h) and the gateway table operations are defined in gateway_table(cc,h). A new header type was also defined in aodv_packet.h to support Gateway Hello-Messages.

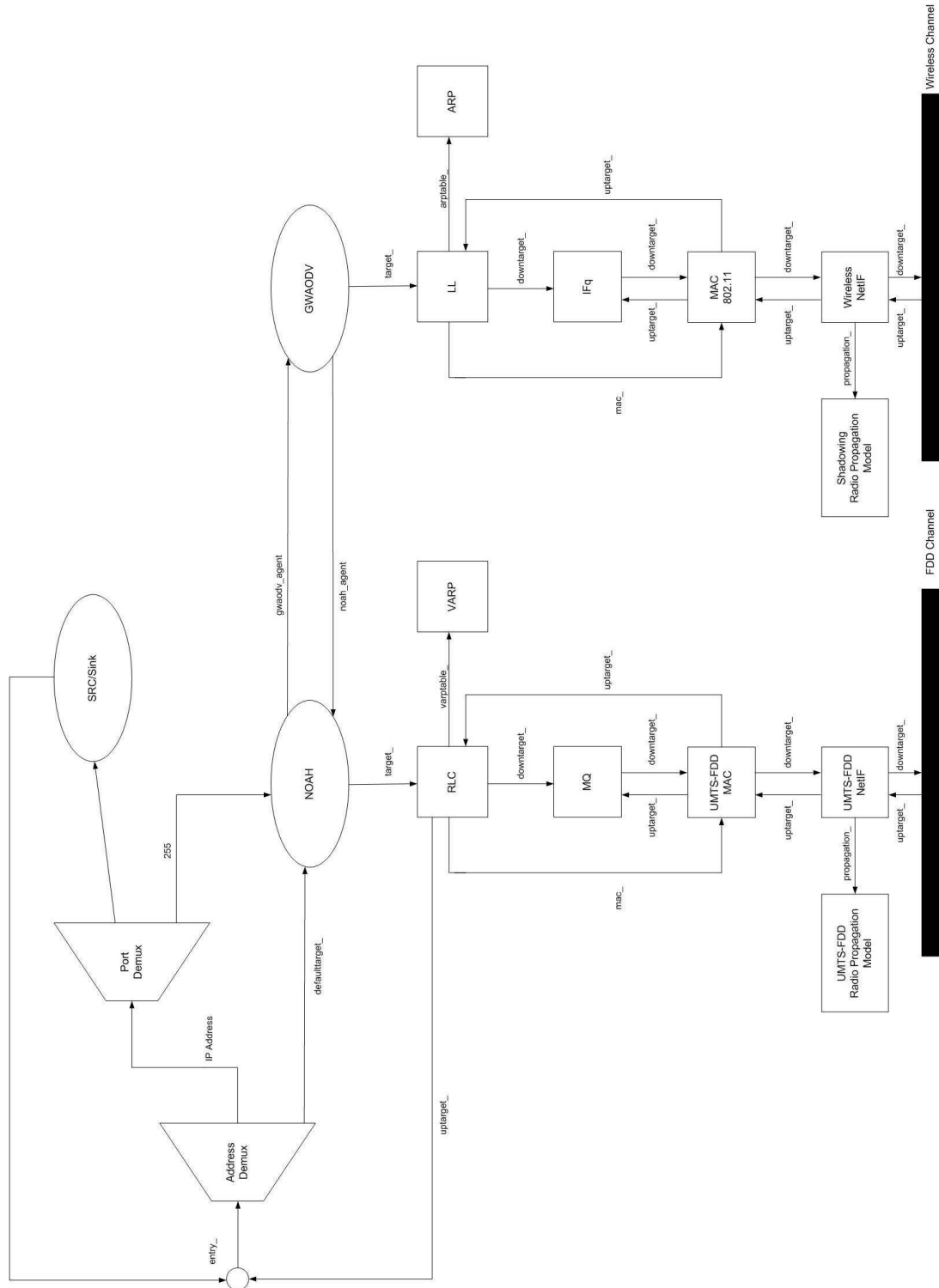


Figure 10.1: Schematic of a double interface wireless node

10.2.3 Modifications to UMTS_MAC_FDD

The `umts_mac_fdd.cc` had to be modified to maintain an associativity level between the NodeB and the UE. This info is then passed to the RLC layer and finally to the NOAH agent.

10.2.4 Modifications to NOAH

This agent (`noah.cc`) had to be modified in order to support the arrival of AODV messages and to decide between using the UMTS interface or the 802.11 interface for data packets depending on the reception or not of a NodeB's signal. The associativity level between the NodeB and the UE is passed from the NOAH agent to the AODV agent to update the node's gateway table.

10.2.5 Other Modifications

Other files such as `rlc.cc` or `channel.cc` have also been modified to adjust to the new additions. As a general rule, any modification is marked by a start label of: `//S ADDED BY BAY` and an end label of `//E ADDED BY BAY`.

Data packets forwarded by a gateway UE from the Ad-Hoc network to the Cellular network bypass the UMTS RLC mechanisms and go straight to the interface queue. The reason for this "work-around" is that a complex mechanism involving packet sequence numbers is implemented by the RLC and if forwarded data packets were to be processed by the RLC as regular packets originating from the UE, the sequence number count would be corrupted. This would then generate two different packets with the same sequence number which would cause malfunctions in the RLC layer protocol.

10.3 Simulation Configurations

The simulations which have been carried out in this study rely on CBR¹ between sources in the Ad-Hoc network and destinations in the wired network. The analyses of the results were focused on the following parameters:

End-to-end delay; the time interval between the moment a packet is sent by the source and the moment it is received by the destination. In the simulations presented below, the end-to-end delay is not entirely accurate; it represents the time interval between the moment the packet is sent and the moment it is received by the RLC layer of the NodeB. However, since the delay on the wired network is minimal compared to the delay measured in the Ad-Hoc network, this approximation does not corrupt the results of the simulations.

¹Constant Bit Rate

Packet delivery fraction; the fraction of received packets to sent packets. The number of sent packets includes messages which have been dropped by GWAODV if no gateway route was available. These messages are generated by the source but never transmitted on the network.

Routing overhead; the amount of control messages sent. The overhead generated by this protocol is solely due to Hello-Messages. Since no modifications were made to the time interval between messages, the calculated overhead would be the same using AODV rather than GWAODV.

To generate traffic, a tool developed by CMU-Monarch was used. `cbrgen.tcl` found in the `"/indep-utils/cmu-scen-gen"` folder of the simulator. This traffic generator takes as parameters: the type of traffic needed, the number of nodes in the simulation, a seed for the random number generator used to determine which nodes establish connections and at what time they start, the maximum number of connections and the packet sending rate. `cbrgen.tcl` was modified into `cbrgen-mod.tcl` in order to create connections between Ad-Hoc and wired nodes instead of "inter-Ad-Hoc" connections.

For node mobility, the `setdest`² utility version 2, developed by U. Michigan and located in the same folder has `cbrgen.tcl`, generated node movement. `setdest`'s parameters are: maximum speed, minimum speed, number of nodes, pause time, pause type (uniform $[0, 2*\text{pause time}]$ or constant), speed type (uniform or normal $[\text{min}, \text{max}]$), maximum time, width of space and height of space.

10.4 General Performance

To analyze the general performance of this protocol, several test on a 50 node topology and with 700s of simulation time, were performed. Every simulation shared the same mobility scenario generated by `setdest`³, with a speed between 1m/s and 2m/s, on the same surface of 1500m by 300m. Three different type of traffic scenario were used, one with a maximum number of connections of 10, a second with 20 and the last one with 40 maximum connections. For each type of traffic scenario, five simulations were run with a varying packet rate. All simulations used the Shadowing propagation model with a path loss of 3.5 and a deviation of 4.

Figure 10.2 gives a summary of `setdest`'s parameters.

²`setdest` implements the Random Way Point algorithm[31]

³`mob-50.tcl`

	Function	Value
-M	maximum speed	2.0m/s
-m	minimum speed	1.0m/s
-n	number of nodes	50
-P	pause type	2 (uniform)
-p	pause time	300s
-s	speed type	1 (uniform)
-t	maximum time	700s
-v	version	2 (UM)
-x	width of space	1500m
-y	height of space	300m

Figure 10.2: Setdest Parameters

10.4.1 Results

Average End-To-End Delay

Per Maximum Connections:

As we can see on figure 10.3 the average end-to-end delay increases as more connections are established. This is due to the increase in the network traffic load; nodes have to process more packets and thus packets spend more time in queues.

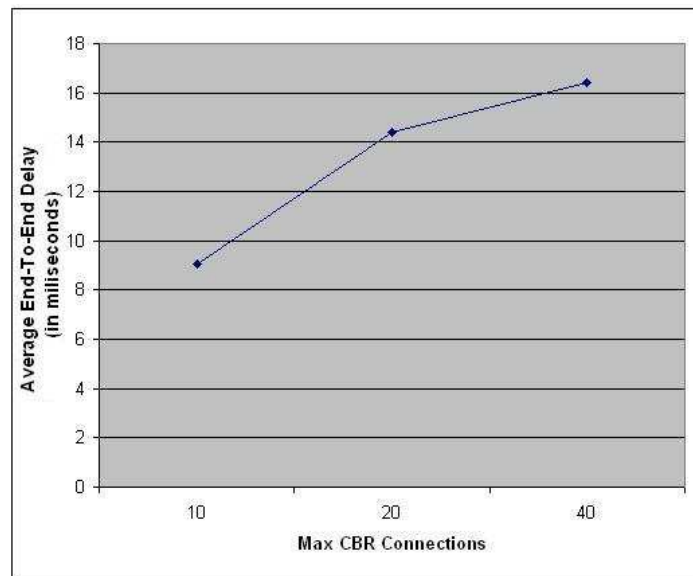


Figure 10.3: Average End-To-End Delay per maximum number of connections

As a function of time:

Figure 10.4 confirms the fact that as more sources begin to transmit data, the average end-to-end delay increases.

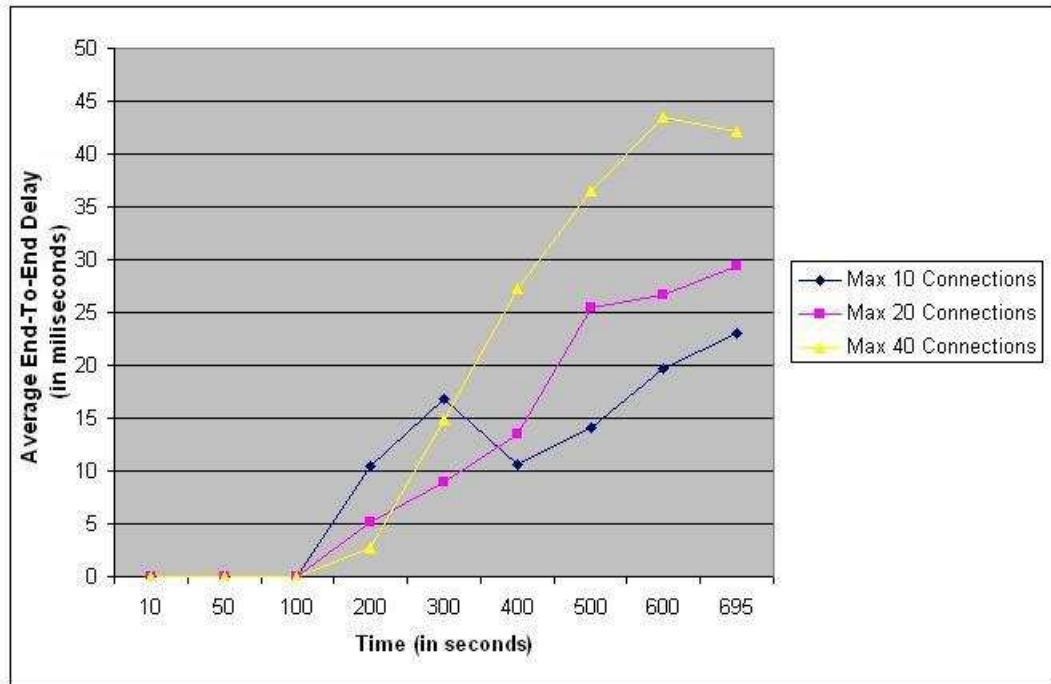


Figure 10.4: Average End-To-End Delay as a function of time

The average end-to-end delay in every simulation is however, well below 150ms which is the maximum delay required for an interactive conversation (voice or multi-media). Above 150ms, voice quality drops below PSTN⁴ quality.

Average Packet Delivery Fraction

In the two figures below, the simulations consisting of maximum 10 connections well surpass the other simulations in terms of average packet delivery ratio. This is due on one part by the fact that, one simulation for maximum 10 connections is extremely prolific and on the other hand, scenarios with maximum 20 connections suffer from two very poor simulations in term of packet delivery ratio. The simulations with maximum 40 connections however, seem to stabilize between 10% and 17%.

A possible explanation for this behaviour is that since the first type of simulation deals with only a maximum of 10 connections, if one or two active

⁴Public Switched Telephone Network

nodes benefit from long lasting stable routes, they will compensate for the loss suffered by less fortunate mobiles. In a scenario with a maximum of 20 connections, if the majority of active nodes are isolated, the amount of packets generated and lost will be too much compared to the successful deliveries. For the scenarios with maximum 40 connections, the chances for all active nodes to be isolated aren't as high as with a maximum of 20 connections and overall a good 10 to 17% of packets are delivered.

The 20 connection scenarios seem to be in a "dead zone" where the number of active nodes is too important for a small number of fortunate nodes to compensate for isolated mobiles and not high enough to prevent a majority of active nodes from being isolated. This however is speculation, further testing is needed to confirm or refute this theory.

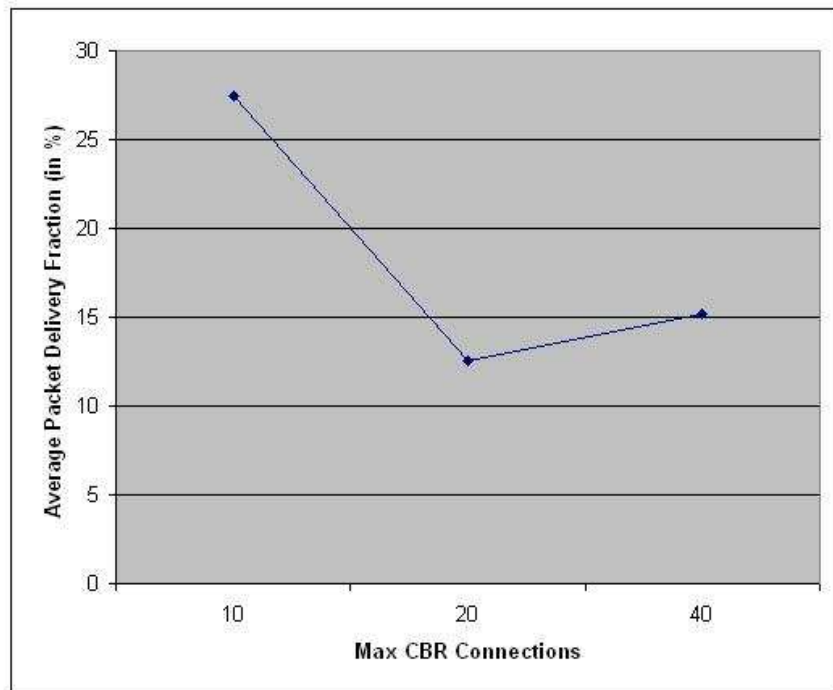


Figure 10.5: Average Packet Delivery Fraction per maximum number of connections

Average Overhead

The average overhead is close to constant. The reason for this is that only Hello and Gateway-Hello control messages are sent in these simulations. There are no RREQ, RREP and RERR because we did not consider "inter-Ad-Hoc" communications.

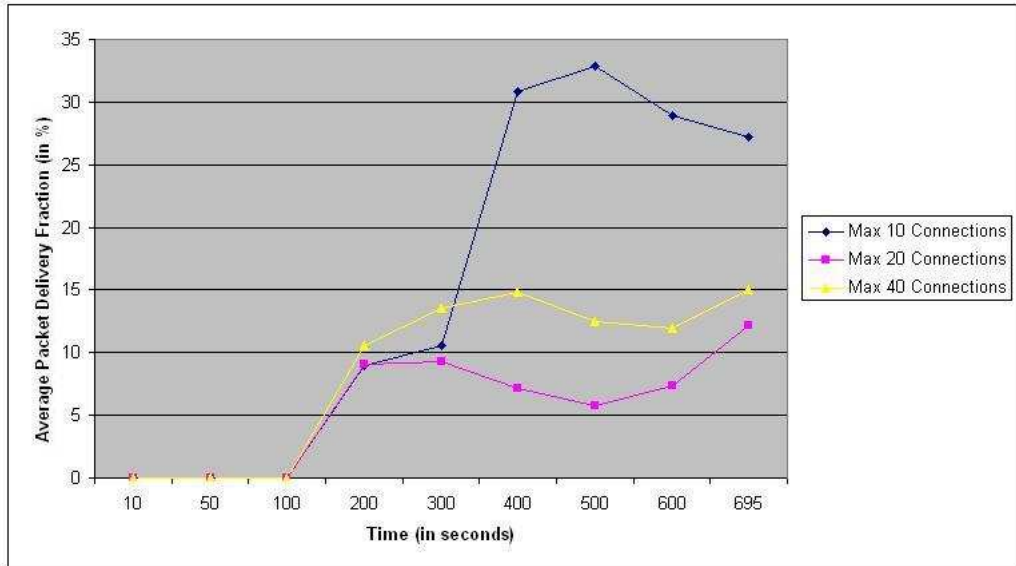


Figure 10.6: Average Packet Delivery Fraction as a function of time

As was said earlier, the generated overhead would have been the same if AODV was used instead of GWAODV for these simulations.

10.5 Different Propagation Models

To fully understand the contrast between an indoor environment and a best case scenario (i.e. a desert plane) with regards to radio propagation, the same simulation configuration using, cbr-40c-3 a traffic scenario, was run using the Freespace propagation model.

An urban scenario was also run on the same traffic file as above but with a different mobility scenario. The maximum speed was raised to 10m/s to account for vehicle movement. The propagation model used is Shadowing but with a path loss exponent and deviation of 3.

The following figures: 10.8, 10.9 and 10.10 illustrate the results of the three simulations.

As expected, the packet delivery fraction is the highest using the Freespace model. However the end-to-end delay is the smallest in Shadowing (3.5, 4.0) and the worst in Freespace. Since less packets are dropped when searching for a gateway (see figure 10.9) using the Freespace model, more packets are processed by nodes due to the existence of several gateway routes. This contributes to the increase of the network traffic load and as we have seen earlier this increases end-to-end delay.

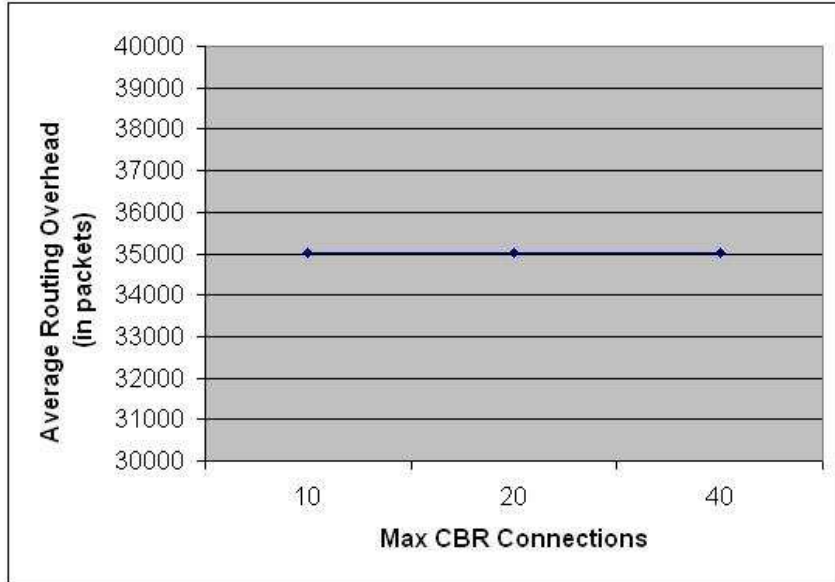


Figure 10.7: Average Overhead per maximum number of connections

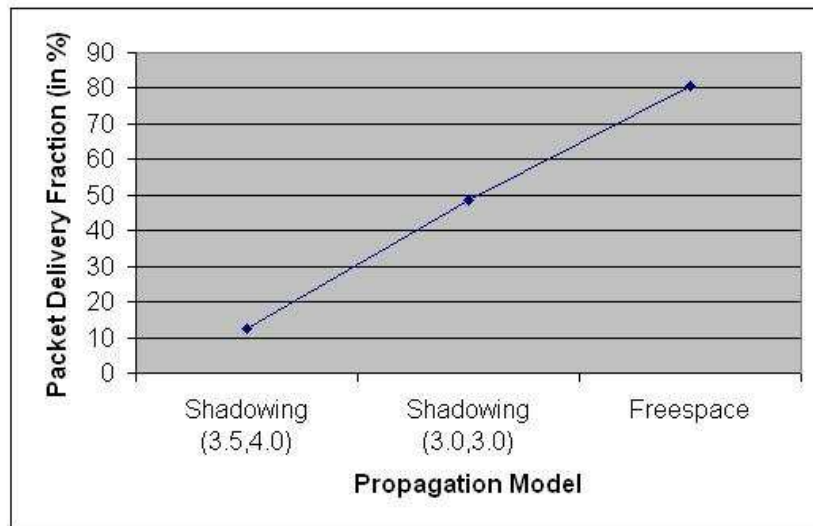


Figure 10.8: Packet Delivery Fraction per Simulation

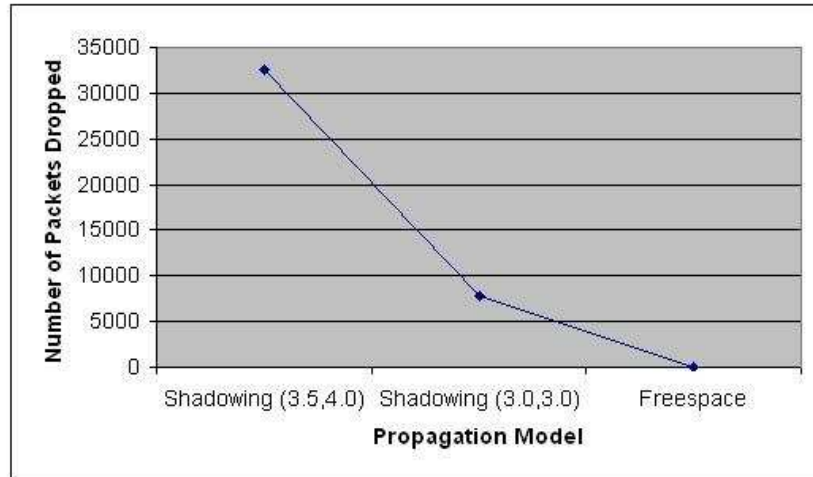


Figure 10.9: Amount of dropped packets due to the lack of gateway routes

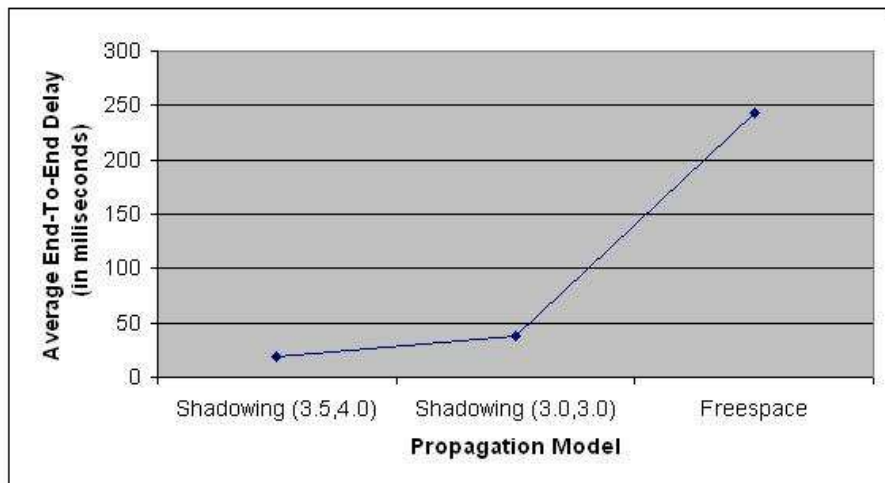


Figure 10.10: Average End-To-End Delay per Simulation

10.6 Varying Packet Rate

This section will study the impact of an increasing packet rate on the end-to-end delay and the packet delivery fraction. Three simulations were run with the same mobility scenario and same propagation model on a 50 node topology and a maximum of 40 connections. A different traffic scenario was used for each simulation; one scenario with a 4 packets/sec rate, a second with a 7 packets/sec rate and a third with a 10 packets/sec rate.

As expected, since the network traffic load increases with a higher packet sending rate, the average end-to-end delay increases but remains under the 150ms mark. The packet delivery fraction decreases since more packets are dropped when no gateway route is present.

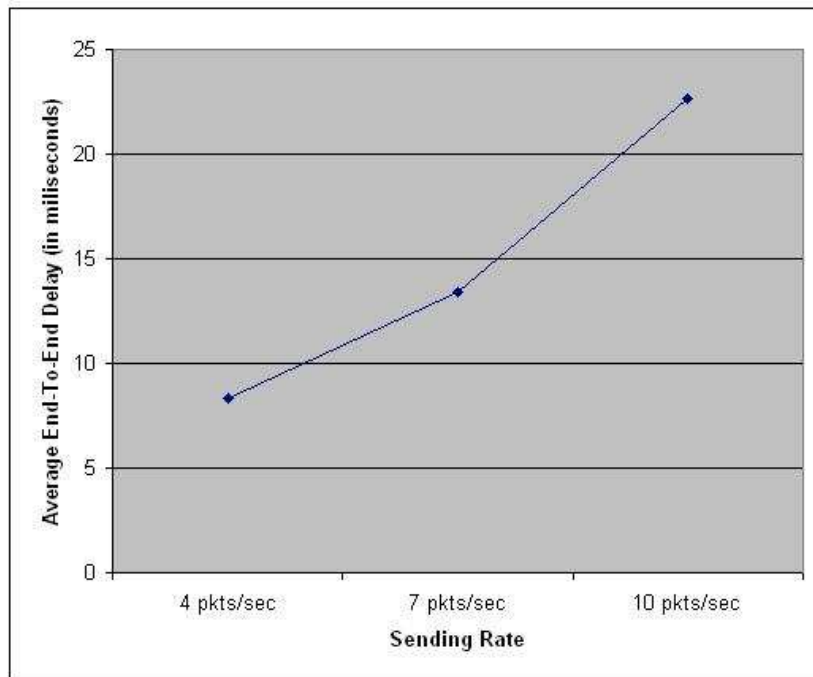


Figure 10.11: Average End-To-End Delay per Sending Rate

10.7 Conclusion

The general performance tests show that this protocol is potentially able to support voice traffic with regards to end-to-end delay. For the packet delivery fraction, if we consider only the packets sent when a gateway route is present in a node's gateway table, then the percentage is extremely high; for instance, the average packet delivery fraction for a scenario with maximum

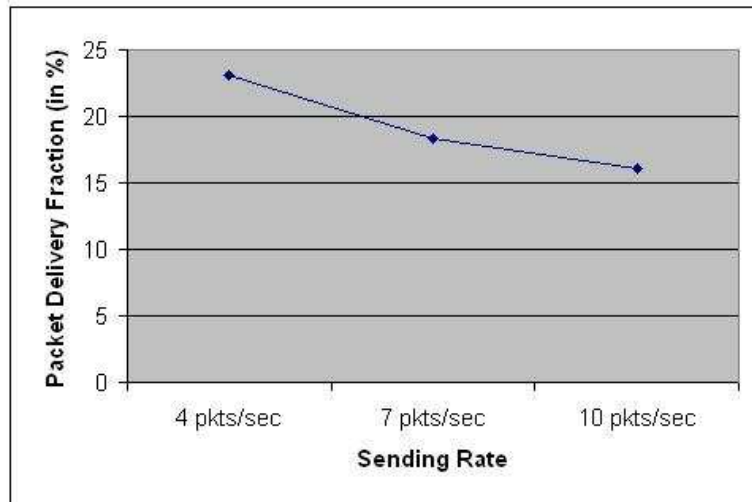


Figure 10.12: Packet Delivery Fraction per Sending Rate

40 connections is 62.7954%. However 10 to 17% in an Ad-Hoc infrastructure is not bad. Nevertheless, the routing overhead, is excessive if the majority of the nodes in the Ad-Hoc network are within reach of a NodeB.

Chapter 11

Conclusion and Future Work

The main objective of this study was to develop an original protocol allowing the extension of a cellular network through an Ad-Hoc network. In order to achieve this, several notions were presented to place this work in its context.

The first chapter of this study focused on the cellular network infrastructure and its evolution from the late 40s to today's recent developments. A detailed description of UMTS was given since the simulations carried out make use of such a network architecture.

Next, the second network infrastructure used in this study was presented: Ad-Hoc networks. This chapter featured a general presentation of MANETs and also an in depth view of the different categories of protocols and their mechanisms using several examples.

Then, the reasons motivating the interconnection of such networks were established along with the work which has already been done in this field. The main factor driving this study is the unpredictable nature of indoor radio signal propagation which can cause "dead spots" preventing devices from gaining access to a cellular network.

The personal contribution part of this study started with a detailed description of AODV and ABR, the two protocols on which are based the developed protocol: GWAODV.

The hypotheses and assumptions/limitations of this study preceded a complete description of GWAODV. This routing protocol's mechanisms were viewed in detail and illustrated through several figures.

Finally, the network simulator, NS-2, was presented along with the modifications needed for this study. The simulation configurations and results ended this work by coming to the conclusion that GWAODV could potentially support voice traffic when considering end-to-end delay and that packet delivery fraction seems to be satisfactory when dealing with Ad-Hoc networks.

However, more testing is needed to fully understand the impact that

this protocol has on these parameters and what effect this would have on real world traffic. For instance, several simulations need to be run using a different metric order when invoking the gateway selection algorithm and analyze the results; the packet delivery fraction may increase if the priority is on associativity.

Bibliography

- [1] Charles E. Perkins and Elizabeth M. Belding-Royer and Samir R. Das. Ad Hoc On-Demand Distance Vector Routing Protocol. In *draft-ietf-manet-aodv-13*, February 2003
- [2] Chai-Keong Toh. Long-lived Ad Hoc Routing based on the Concept of Associativity. In *draft-ietf-manet-longlived-adhoc-routing-00.txt*
- [3] The ns Manual (formerly ns Notes and Documentation). In *<http://www.isi.edu/nsnam/ns/>*
- [4] Andrew S. Tanenbaum. Third-Generation Mobile Phones: Digital Voice and Data. In *Computer Networks*, fourth edition, page 166. Prentice Hall PTR, 2003
- [5] Gislain Bocq. UMTS Services. In *Chapter: UMTS*, edition 1.6, page 8, 2006
- [6] The 3rd Generation Partnership Project Agreement. December 1998.
- [7] 3rd Generation Partnership Project. Feasibility study on 3GPP system to Wireless Local Area Network (WLAN) interworking (Release 6). In *Technical Specification Group Services and System Aspects*, October 2003
- [8] IEEE 802.11. In *http://en.wikipedia.org/wiki/IEEE_802.11*
- [9] Charles E. Perkins and Pravin Bhagwat. DSDV Routing over a Multihop Wireless Network of Mobile Computers. In *Ad Hoc Networking, Addison-Wesley*, chapter 3, pages 53-74, 2001
- [10] Thomas Clausen and Philippe Jacquet. Optimized Link State Routing Protocol (OLSR). In *RFC3626*, October 2003
- [11] David B. Johnson and David A. Maltz and Yih-Chun Hu. The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks (DSR) In *Tomasz Imielinski and Hank Korth*, Mobile Computing, volume 353, pages 153-181. Kluwer Academic Publishers, 1996. Chapter 5.

BIBLIOGRAPHY

- [12] Vincent D. Park and M. Scott Corson. A Highly Adaptive Distributed Routing Algorithm for Mobile Wireless Networks. At *IEEE Conference on Computer Communications*, INFOCOM'97, April 7-11, 1997, Kobe, Japan
- [13] Zygmunt J. Haas and Marc R. Pearlman. The Zone Routing Protocol (ZRP) for Ad Hoc Networks. In *haas-draft-ietf-manet-zone-zrp-00*, November 1997
- [14] Andrew S. Tanenbaum. Radio and Microwave Transmission. In *Computer Networks*, fourth edition, page 103-105. Prentice Hall PTR, 2003
- [15] An Introduction To Wireless Communications.
http://gk12.harvard.edu/modules/Radio_Propagation.doc
- [16] Gislain Bocq. Les trajets multiples. In *Chapter: Concepts de Propagation*, edition 2.6, page 4, 2006
- [17] Theodore S. Rappaport. Wireless Communications - Principles & Practice, IEEE Press, 1996, page 130.
- [18] The ns Manual (formerly ns Notes and Documentation). In <http://www.isi.edu/nsnam/ns/> page 186.
- [19] The ns Manual (formerly ns Notes and Documentation). In <http://www.isi.edu/nsnam/ns/> page 188.
- [20] H. T. Friis. A note on a simple transmission formula. In *Proc. IRE*, 34, 1946.
- [21] Hongyi Wu, Chunming Qiao, Swades De, Ozan Tonguz. Integrated Cellular and Ad-Hoc Relaying Systems: iCAR. In *IEEE on Selected Areas in Communications*, Vol. 19, No. 10: 2105-2115, October 2001.
- [22] H. Luo, R. Ramjee, P. Sinha, L.E. Li and S. Lu. UCAN: A unified cellular and Ad-Hoc network architecture. In *Proc. MobiCom*, San Diego, CA, USA, 2003.
- [23] Lorena Senador-Gomez Lazaro. Aspects For The Integration Of Ad-Hoc and Cellular Networks In Indoor Environments. At *Universite Libre de Bruxelles*, 2006
- [24] Anand A. Janefalkar, Kaushik Josiam, Dinesh Rajan. Cellular Ad-hoc Relay for Emergencies.
- [25] Bharat Bhargava, Xiaoxin Wu, Yi Lu, Weichao Wang. Integrating Heterogeneous Wireless Technologies: A Cellular Aided Mobile Ad hoc Network (CAMA). *Mob. Netw. Appl.* 9, 4 (Aug. 2004), 393-408

BIBLIOGRAPHY

- [26] Jean-Pierre Hubaux, L. Buttyan and S. Capkun. The Quest for Security in Mobile Ad Hoc Networks. In *Proceedings of the 2nd ACM international Symposium on Mobile Ad Hoc Networking & Computing* (Long Beach, CA, USA, October 04 - 05, 2001). MobiHoc '01. ACM Press, New York, NY, 146-155.
- [27] Naouel Ben Salem, Levente Buttyan, Jean-Pierre Hubaux, Markus Jakobsson. A Charging and Rewarding Scheme for Packet Forwarding in Multi-hop Cellular Networks. In *Proceedings of the 4th ACM international Symposium on Mobile Ad Hoc Networking & Computing* (Annapolis, Maryland, USA, June 01 - 03, 2003). MobiHoc '03. ACM Press, New York, NY, 13-24.
- [28] Mohan, M. and Joiner, L. L. Solving billing issues in ad hoc networks. In *Proceedings of the 42nd Annual Southeast Regional Conference* (Huntsville, Alabama, April 02 - 03, 2004). ACM-SE 42. ACM Press, New York, NY, 31-36.
- [29] Standard Deviation. In http://en.wikipedia.org/wiki/Standard_deviation
- [30] Andrew S. Tanenbaum. Reference Models. In *Computer Networks*, fourth edition, page 49. Prentice Hall PTR, 2003
- [31] The Random Waypoint Model. In <http://www.netlab.tkk.fi/esa/java/rwp/rwp-model.shtml>