ACO-BASED ROUTING ALGORITHMS FOR WIRELESS MESH NETWORKS

Chen Yuan

A Thesis

In

The Department

Of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Applied Science at
Concordia University
Montréal, Québec, Canada
November 2009

Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-67294-5
Our file Notre référence
ISBN: 978-0-494-67294-5

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

ABSTRACT

ACO-Based Routing Algorithms for Wireless Mesh Networks

Chen Yuan

The popularity of Wireless Mesh Networks (WMNs) is growing exponentially in recent years, due to their flexible deployment and compatible communication features. As a key technology for next-generation wireless networking, WMNs promise an attractive future to both academic and industrial world. However, current WMNs are short in optimal routing protocols. Instead, many WMNs use the routing algorithms from ad hoc networks, which have different network features. Thus, routing becomes the most urgent issue that needs to be solved.

In this thesis, routing problems in WMNs are discussed in different aspects, and then several proposed solutions in state-of-the-art are introduced with their advantages and disadvantages. Ant-In-Mesh routing protocol and the enhanced version are proposed for WMNs, inspired by traditional Ant Colony Optimization (ACO) algorithm, to deal with new challenging characters of WMNs. Periodical Mesh update is performed between neighbors, to keep the network alive. With these updated information at all the hosts, various Ants can collect the fresh routing data while they are launched for different purposes, also, the per-hop and end-to-end routing metrics can be calculated. Upon new connection requests, route discovery is carried out. After the routes are set up, proactive route maintenance is performed on each route. Several popular routing protocols and our algorithms are simulated and compared using Qualnet. The simulation results show that our algorithms outperform the others, in terms of packet delivery ratio and end-to-end delay, as the mobility and network size increase.

ACKNOWLEDGMENTS

First, I would like to express my sincere gratitude to my research supervisor, Dr. Anjali Agarwal, for her invaluable guidance, advice, friendship, and financial support during my entire Master's program.

Also, I would like to thank all the course professors, Drs. M. Omair Ahmad, Dongyu Qiu, Amir G. Aghdam, and Chadi Assi, from whom I gained precious knowledge and widened my professional view of the field. Moreover, I would like to thank my colleagues in the research group, Ms. Saloni Neri, Mr. Ahed Al-Shanyour and Mr. Mahmudur Rahman, for all the helpful discussions and sharing.

Last, but not least, I am very grateful towards my parents, whose constant support and love made it possible to finish my work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

Access Point	AP
Ad hoc On-Demand Distance Vector	AODV
Ant Colony Optimization	ACO
Ants-in-Mesh	AIM
Automatic Repeat reQuest	ARQ
Basic Service Set	BSS
Carrier Sense Multiple Access with Collision Avoidance	CSMA/CA
Contention Free Period	CFP
Destination-Sequenced Distance Vector	DSDV
Direct Sequence Spread Spectrum	DSSS
Distributed Coordination Function	DCF
Distributed Spanning Trees based routing protocol	DST
Distribution System	DS
Dynamic Source Routing	DSR
Expected Transmission Count	ETX
Expected Transmission Time	ETT
Forward Error Correction	FEC
Frequency Hopping Spread Spectrum	FHSS
Global Positioning System	GPS
Global State Routing	GSR
Hybrid Wireless Mesh Protocol	HWMP
Light-weight mobile routing	LMR
Link-State-Advertisements	LSAs
Load-aware-Ant-In-Mesh	LAIM
Local Area Network	LAN
Logical Link Control	LLC
Media Access Control	MAC
Mobile Ad Hoc Network	MANET
Multi Point Replaying	MPR
Multi-Radio Link Quality Source Routing	MR-LQSR
Multi-radio Unification Protocol	MUP
Network Allocation Vector	NAV
Open System Interconnection	OSI

Optimized Link State Routing	OLSR
Orthogonal Frequency Division Multiplexing	OFDM
Packet Delivery Ratio	PDR
Personal digital assistant	PDA
Point Coordination Function	PCF
Receiver Based Packet Pair	RBPP
Receiver Only Packet Pair	ROPP
Round Trip Time	RTT
Sender Based Packet Pair	SBPP
Transmission Queue Length	TQL
Wireless Local Area Network	WLAN
Wireless Mesh Network	WMN
Wireless-Fidelity	Wi-Fi
Worldwide inter-operability for Microwave Access	WiMAX
Zone Routing Protocol	ZRP

CHAPTER 1

Introduction

This chapter presents the background and development of wireless networking, the important IEEE 802.11 standards, and the characters of wireless mesh networks.

In modern world, wireless networking becomes more and more significant in human's life. With wireless connection, users can get rid of the location constraint and annoying wire lines, thus people can communicate with the others and get fresh information no matter where they are. Currently, mobile-enable devices are occupying the entire electronic market, most consumer devices are now equipped with Wireless-Fidelity (Wi-Fi), Worldwide inter-operability for Microwave Access (WiMAX), Infrared, or Bluetooth technology, such as cellular phone, laptop, Personal digital assistant (PDA), and even the digital camera and printer.

Besides these consumer devices, wireless networking also influences the classic industry. People would consider to power up their cars and houses by installing radio or Bluetooth sensors or transmitters at an affordable cost. Most highways or roads now have on-line video, in order to monitor the public safety and allocate the traffic fairly. Furthermore, spontaneous (Emergency/Disaster) networking also benefits from wireless technology, under certain circumstances, where human can hardly reach or under time-constraint environments, wireless networking is fundamental in assisting people to deal with an urgent situation. It can be imagined that, in the future, people all over the world can enjoy the convenience that this interesting technology offers.

1.1. Wireless Network Background

Within recent years, Wireless Local Area Networks (WLANs) have rocketed the quality of human's life; they entirely stimulated a powerful revolution on how people communicate with each other, and how the latest news can spread over the world. As we all know, Internet service has become indispensable in modern life, it is like a super huge data center with all the information we need. With internet, we can chat with family, friends, and colleagues in the other end of the earth; international video meetings that used to exist only in the dreams now occur every moment; we can do business at home, without even handing out the business card on the street; all these fancy things now come true.

Later on, how to access the Internet resource at any place anytime becomes the public concern, and this rising demand spurred the development of various wireless networking technology. Currently, public WLANs are available at many cafeterias, libraries, hospitals, airports and universities, which are called Wi-Fi hotspots. These Wi-Fi hotspots usually have one or more access points which are pre-connected to the available wired backbone network, so deploying new Wi-Fi hotspots needs some extensive infrastructure and considerate networking plans to manage the budget.

1.1.1 IEEE 802.11 Wireless Standards

In 1997, IEEE first proposed the 802.11 standard for wireless networking, which defines a set of Wireless LAN standards developed by the working group 11 in IEEE standards committee. From then on, all the vendors have the same rules to follow, and the mobile users can communicate with each other at designed speeds.

According to IEEE standard definition, 802.11 groups define the specifications of Physical and Media Access Control (MAC) Layer, as shown in Fig. 1-1.

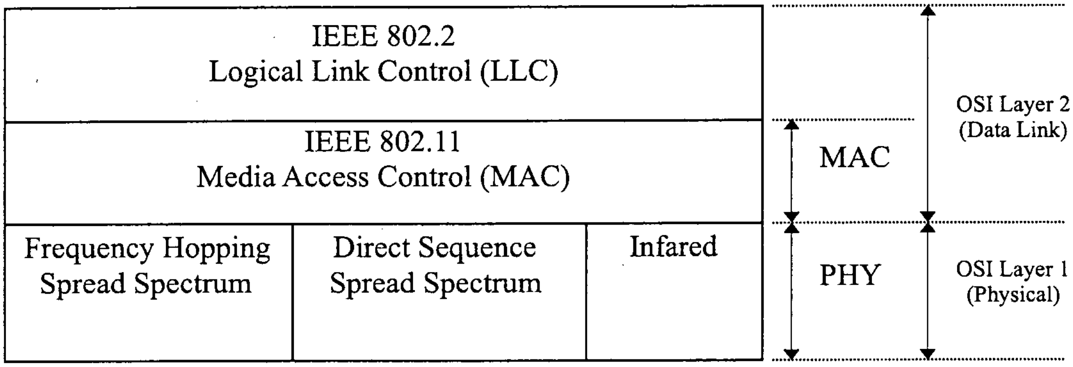


Fig. 1-1 IEEE 802.11 Physical and MAC Layer specifications

In IEEE 802.11, there are basically three types of modulation schemes: Frequency Hopping Spread Spectrum (FHSS), Direct Sequence Spread Spectrum (DSSS), and Orthogonal Frequency Division Multiplexing (OFDM). FHSS and DSSS operate at 2.4GHz band, both of them support data rates of 1 and 2 Mbps, and DSSS can also support data rates up to 11Mbps; and OFDM normally utilizes the band of 5GHz, while providing data rate up to 54Mbps in 802.11a. In the following part, we introduce the 802.11 family:

1. IEEE 802.11: It is the first proposed IEEE wireless standard, which was named after the whole family of wireless standards, and now it is often referred to 802.11 Legacy. It came out in 1997 and operates at 2.4GHz, with FHSS or DSSS modulation scheme, it provides speeds between 1Mbps and 2Mbps.
2. IEEE 802.11a: In 1999, this popular standard was brought into the big family, with maximum 54Mbps data rate, using OFDM modulation algorithm. The 802.11a standard provides 8 radio channels in the 5 GHz frequency band.
3. IEEE 802.11b: It is currently the most widely used one, which was a marvelous

success when first ratified in 1999. This standard offers a maximum throughput of 11 Mbps and a communication range up to 300 meters in the open air, using DSSS modulation technique. It operates in the 2.4 GHz band, with 3 radio channels available for use.

4. IEEE 802.11g: This standard offers the same speed as 802.11a up to 54 Mbps, while operating on the 2.4 GHz frequency range. It adopts DSSS when the speed is less than 20Mbps, and OFDM for speeds greater than that. The 802.11g is a special standard, which is also compatible with the 802.11b, which means it can support some old devices using 802.11b.
5. IEEE 802.11s: It is a draft IEEE 802.11 amendment especially made for mesh networking, defining how various wireless devices can interconnect to create an ad-hoc network. In 802.11s, legacy 802.11 MAC standard is extended by defining an architecture and protocol that support both broadcast/multicast and unicast delivery, using "radio-aware metrics". Furthermore, 802.11s also includes a new mandatory routing protocol named Hybrid Wireless Mesh Protocol (HWMP) [1].
6. IEEE 802.11n-2009: It is an amendment to the IEEE 802.11-2007 wireless networking standard to improve network throughput. Multiple-input multiple-output (MIMO) technology is added into this standard and 40 MHz channels are adopted at the physical (PHY) layer, instead of 20 MHz. Moreover, the maximum raw OSI physical layer (PHY) data rate is increased from 54 Mbit/s to 600 Mbit/s with the use of 4 spatial streams at a channel width of 40 MHz, which doubles the channel capacity.

In 802.11 MAC layer, media access can be expressed in two forms: distributed coordination function (DCF) and point coordination function (PCF). Normally, all the wireless stations (STAs) are required to implement DCF, as it is the fundamental way to support media access control, based on carrier sense multiple access with collision avoidance (CSMA/CA) protocol. This algorithm avoids collision at the STAs, by only enabling packet transmission when the STAs do not receive any message from the others. However, this mechanism is implemented in the PHY layer, which is not enough. Therefore, IEEE 802.11 adds in a virtual carrier sense mechanism, which is based on exchange of Request To Send (RTS) and Clear To Send (CTS) frames. RTS/CTS communication scheme enables the stations to hold the medium channel for a specified period of time, to better avoid collision or retransmission. In order to understand how RTS/CTS works, a brief introduction to 802.11 MAC frame structure is necessary. There are three types of frames: Control, Management, and Data. The frame format of 802.11 MAC is shown in Fig. 1-2.

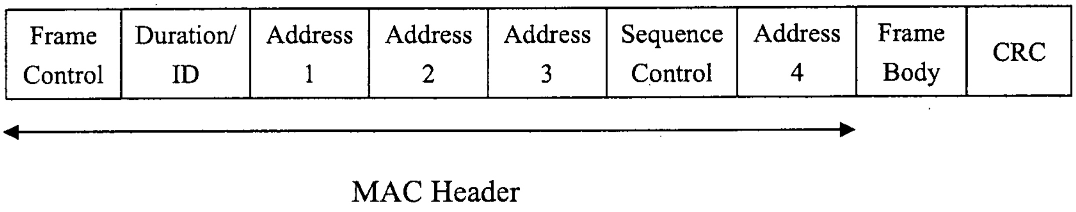


Fig. 1-2 802.11 MAC frame format

Frame Control: It contains many flags and status information introduced below:

- Protocol Version: This field consists of 2 bits, which can be used for future extended version of protocol standard. Currently, it is fixed as 0.
- Type and Subtype: This 6-bit field represents the type of the frame, including management, control and data, with different combination value.

- ToDS: When the frame is addressed to the AP to be forwarded to the Distribution System (DS), this bit is set to 1, otherwise, it is set to 0.
- FromDS: This bit is set to 1 if the frame is coming from the Distribution System.
- More Fragments: When there are more outstanding fragments belonging to the same frame following this received fragment, this bit is set to 1.
- Retry: This is the flag of fragment retransmission.
- Power Management: This bit indicates the next power management mode of the station, after the transmission of this frame.
- More Data: This bit is used by the Access Point (AP) to indicate whether there are more frames buffered to this station, which can continue polling or change power mode based on this information.
- WEP: This bit is the flag of WEP algorithm, which encrypt the frame body.
- Order: This bit indicates that this frame is strictly ordered.

The detailed format is shown in Fig. 1-3.

Item	Protocol Version	Type	Subtype	To DS	From DS	More Fragment	Retry	Power Management	More Data	WEP	Order
Bits	2	2	4	1	1	1	1	1	1	1	1

Fig. 1-3 Details of Frame Control field

Duration/ID: This field has two meanings depending on the frame type. If it is power-save poll message, this field presents the Station ID; otherwise, this is the duration value used for the Network Allocation Vector (NAV) calculation.

Address Fields: Any frame may contain up to 4 addresses depending on the values of "ToDS" and "FromDS", which are defined in the Control field. Address 1 is always the recipient address, if "ToDS" is set, then this is the address of the AP, otherwise it is the address of the end-station; Address 2 is always the transmitter address, if "FromDS" is set, then this is the address of the AP, otherwise it is the end-station address. Regularly, Address 3 is the remaining or the missing address. On a frame with "FromDS" set to 1, the address 3 is the original source address, it is the destination address if "ToDS" is set on the frame; Address 4 is only used in Wireless Distribution System, in which the frame is transmitted from one AP to the other.

Sequence Control: This field is used to indicate the order of different fragments from the same frame, it can also be used to recognize packet duplications, since sequence control field contains two subfields: fragment number and sequence number. CRC is short for Cyclic Redundancy Check, which consists of 32 bits.

When a sender STA has something to transmit, it first sends out a RTS frame. The field of Duration/ID in the MAC header is set to the transmission time for transmitting the pending data or management frame, plus one CTS and one ACK frame, plus three Short Interframe Space (SIFS) intervals. The whole time value is called Network Allocation Vector (NAV), which is also the time the undergoing transmission needs to keep the medium. At the receiver STA, if the RTS is received successfully, and the channel is idle, it will send back a CTS frame. Receiving the CTS frame correctly, the sender STA transmits the data payload, waiting for an ACK frame from the receiver if the data transmission is a success. Meanwhile, all the other STAs know that there is a transmission within the medium, as they all hear the CTS frame, so they need to back off for NAV microseconds before the contention period of the next transmission.

On the other hand, PCF is an optional access method, with additional specific support for the transmission of some time-sensitive streams like voice and video data. However, the PCF only makes sense when the network is working in the infrastructure mode. During the contention free period (CFP), the channel access is granted to STAs by polling the STA. If a Basic Service Set (BSS) is set up enabling PCF, time is spliced between the CFP with PCF mode and the Contention Period in the DCF mode.

1.1.2 Wireless Network Architecture

Wireless networks are normally setup according to the Star topology, which is a kind of infrastructure mode. It is easily distinguished from the others, by the presence of an access point, which has entire control of the network and all stations can communicate with each other or the backbone only via this access point. With this access point connecting to the internet, users within the transmission range can access the internet and communicate with each other. However, if the destination station is outside this infrastructure, the message is first sent to the access point, which will further send it out to the destined infrastructure access point, then to the final station.

Moreover, the 802.11 standard also supports another popular Ad hoc mode, in which there is no access point taking charge. This mode is quite convenient to establish, especially in a certain area, where a group of users want to bind together fast to exchange information, like conference or gaming room. All the stations within the range can communicate with each other freely and directly. Due to its topology and characters, it is impossible to access internet only within the ad hoc network. On the other hand, if a source wants to reach some destination station out of the range, certain intermediate point is necessary to form the bridge. Therefore, in a large wireless multi-hop network, the decision that chooses the next hop influences seriously the performance of a routing protocol.

In recent years, wireless mesh network (WMN) architecture has been proposed to fit in the design of last mile telecommunication, also to be a competitive option for increasing broadband Internet penetration. The details of WMN are discussed in the following section.

1.2. Wireless Mesh Network

Recently, wireless mesh networking, a new broadband Internet access technology, draw significant attention from both academic and industrial world. It can be applied to enormous number of environments, e.g., broadband home networking, community and neighborhood networks, enterprise networking, building automation, metropolitan area networks, transportation and emergency systems.

Based on all the precious work of 802.11 WLAN and wired networks, WMN architecture benefits from previous experience from both academe and industry. Normally, WMN is established by some static mesh routers, which are connected to each other by wireless links and also to a few gateway nodes. These gateway nodes may have stable wired connectivity to the Internet.

As introduced in [2], WMNs are self-organized and self-configured, with the radio nodes automatically establishing an ad hoc network and maintaining the mesh connectivity. WMNs usually consist of mesh clients and mesh routers, the mesh clients are often laptops, cell phones and other wireless devices while the mesh routers could be the static and powerful routers. In WMNs, the coverage area of the radio nodes working as a single network is called a mesh cloud, access to this mesh cloud is dependent on the radio nodes working in harmony with each other to create a radio network. A wireless mesh network is reliable and offers redundancy, when one node can no longer operate, the rest of the nodes can still communicate with each other, directly or through one or more intermediate nodes. Moreover, WMNs can also be implemented with various wireless technology including 802.11, 802.16, cellular technologies or combinations of more than one type. Thanks to wireless mesh networking, people can enjoy living in remote areas and small businesses can be operated in rural neighborhoods at affordable costs.

A typical wireless mesh network is shown in Fig. 1-4, in which we can see that the backbone mesh routers can support mesh clients and also other types of networks.

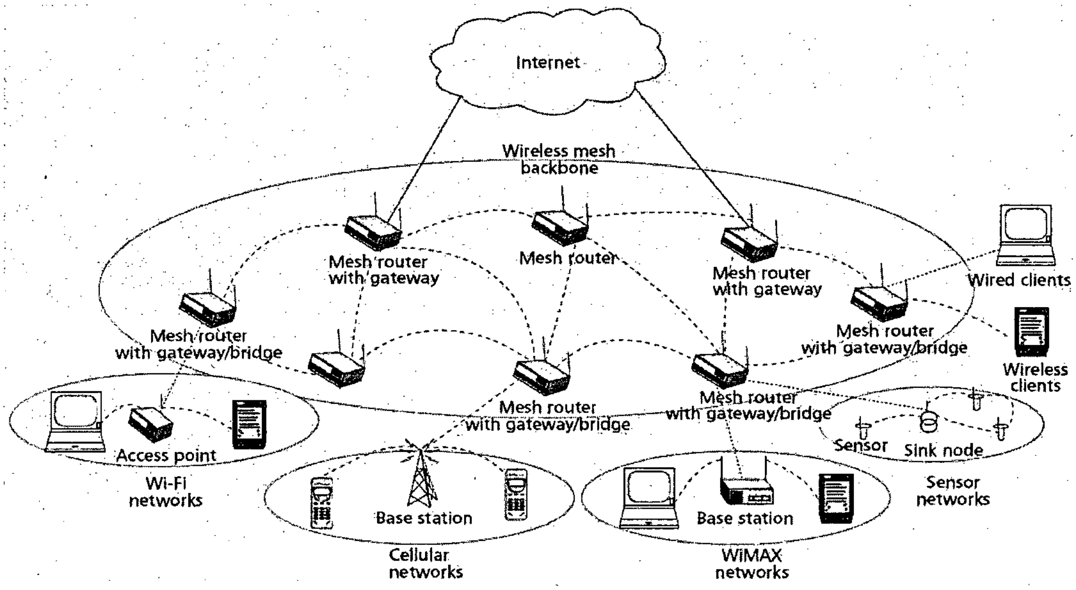


Fig. 1-4 Typical wireless mesh network

Because of WMN's advantageous features, a potential preferential solution is provided to integrate various wireless networks with different standards together, such as cellular network, wireless sensor network, Wi-Fi, and WiMAX networks. Therefore, users from different existing networks can benefit from the entire combination of possible services. Besides all the goodness of wireless networking, WMN provides new advantages such as:

- Internet gateways are less utilized,
- Special backup technology enhances data security,
- The routes are configured dynamically and flexibly,
- Power issues become easy jobs due to lower energy requirement,
- Integration: Mesh hardware is typically small, noiseless, and easily encapsulated in weatherproof boxes. This means it also integrates nicely outdoors as well as in human housing,
- Reliability in WMNs is increased, the network nodes are connected to provide backups: The character of mesh topology and ad-hoc routing promises greater stability in the face of changing conditions or failure at single nodes, which will quite likely be under rough and experimental conditions,
- Ease and simplicity: setup is extremely simple for a box that is pre-installed with wireless mesh software and that uses standard wireless protocols such as 802.11b/g. Since routes are configured dynamically, it is often enough to simply add the box into the network, and attach whatever antennas are required for it to reach one or more existing neighboring nodes.

It is firmly believed that, WMNs will become the most widely utilized network in the near future. Thanks to the efforts of many vendors, WMNs have been implemented in various technology fields, such as automobile, gaming entertainment, and automatic building communication. Moreover, WMNs may be applied to the highways and roads, as swarms of cars equipped with this technology can function as mobile nodes.

1.3. Motivation

WMN promises an attractive future for both academic and industrial world. Thus, numerous researchers coming from various fields, with different backgrounds, are convinced to put their enthusiasm and precious time on research of WMN. On the other hand, a wide range of researchers begin to review the architecture of relatively mature networks, such as Mobile ad hoc network (MANET), wireless sensor network, and IEEE 802.11 networks, from the design view of WMNs.

However, in order to make WMN an ideal choice for wireless networking, we still need constant efforts on research. For example, network performance becomes more and more disappointing as the mobility of network increases; the available protocols for different layers do not consider the variety of input traffic in WMN; the adopted protocols of layer-2 and layer-3 do not scale well in WMN; as the network size grows, the throughput and delay degrade and become out of control [3].

As the history tells us, any new technology must experience the tough developing process to become mature for use. Currently, wireless mesh networking is going through such a period with many demands and challenges, especially in designing the routing protocols, which play a significant role in consummating the wireless technology. Unfortunately, for lack of available routing protocols at both layer-2 and layer-3, WMN designers normally adopt the routing protocols from the other networks, like MANET and wireless sensor network, which have their own characters.

For example, MANET consists of a group of users without any infrastructure, the aim of each user is to reach the others and make peer to peer transmission, and there is no station that takes charge of issues, such as security and authentication service. As introduced in Section 1.2, a wireless mesh network provides more powerful mesh routers, also the access to wired network like internet. All the mesh stations are dedicated to the operation of the whole mesh network, which requires reasonable network management, and thus more knowledgeable routing protocol is needed in WMN. Overall, MANET or sensor networks could be seen as a subset of the WMN. Due to the differences between WMN and other networks, in WMN, utility of routing protocols from other networks like MANET could restrict the development of WMN. Consequently, the possible advantages of WMN cannot successfully emerge, unless suitable routing protocols are proposed.

According to the characteristics of WMNs described in last section, in order to design a suitable routing protocol for WMNs, we need to consider not only the fast adapted and fair load-balancing routing, but also the mature integration of various network metrics. However, most routing protocols proposed for WMNs do not meet all these requirements. Therefore, we aim to design routing protocols based on the requirements. The first step is to formulate a routing metric, which consists of necessary network components, such as the estimated bandwidth, delay, and loss ratio. Moreover, the metric should contain possible parameters that can reflect network load situation. And then, we need to implement this metric into certain optimal algorithm, which rules what and how the mesh stations communicate with each other, and how to route the data.

At present, a number of researchers are making great efforts to achieve the goal of effective and optimal routing [9, 13, and 14]. Among them, there is a branch which obtained the inspiration from the behavior of natural ants. The scholars have observed that ants from the same colony can head to a source of food over the shortest paths, such a talent owes to a chemical substance called pheromone. When ants move between the nest and the source of food, they deposit pheromone, which gives further ants guidance on the choice of route [4]. From numerous experiments, researchers found that Ants' behaviors are quite suitable to design routing protocols in wireless network [26, 27, 28, 29, 30, 31, 32], and in related works, routing protocols using Ant Colony Optimization (ACO) obtained better performance than traditional protocols for wireless network.

From another point of view, a WMN can be seen as a huge Ant system with lots of powerful ant colonies and different food destinations. Instead of finding the shortest paths between the colony and the destinations, we seek for optimal routing paths with consideration of link quality and load situation. In the last decades, many routing protocols were proposed for wired and wireless network, and the related works are analyzed in Section 2.6.

In this thesis, we propose routing metrics that combine the possible estimated network components, delay, bandwidth, loss ratio, and transmission queue length of each station, which can partially reflect the network load situation. They are formulated according to certain weight, since different applications desire different performance on these components. Therefore, the metrics meet the WMNs' requirements and can be applied to any type of routing protocols. Due to the fast reaction and information update of ACO algorithms, we apply the metrics into the routing protocol based on ACO.

1.4. Organization

Chapter two gives a literature review of routing algorithm design and related works, and the ACO algorithm and proposed protocols are introduced, with their advantages and drawbacks. Chapter three presents our first proposed work, Ant-In-Mesh (AIM) routing protocol, which is based on ACO algorithm. It contains on-demand path exploration and proactive path maintenance. We compare AIM with AODV in Qualnet simulator, and show the simulation results accordingly. Chapter four gives the details of Load-aware-Ant-In-Mesh (LAIM), an enhanced routing protocol of AIM, with more accurate estimation scheme for network components and more powerful routing metrics. LAIM is also compared with AODV and AntNet in Qualnet, and the simulation results are shown. In Chapter five, we conclude the thesis and present possible future work.

CHAPTER 2

Routing and Related Works

In this chapter, existing routing algorithms are reviewed from various aspects. Besides, other contents in the design of a routing protocol are introduced in this chapter.

Routing is at the core of the whole network control system. Routing, in conjunction with the admission, flow, and congestion control components, determines the overall network performance in terms of both quality and quantity of delivered service [5]. All the tasks of routing are completed with the help of suitable routing protocols, which make the rules how packets are routed from the source to the destination. Thus, the network performance and fairness are mainly determined by the routing protocol.

2.1. Categories of Routing Algorithm

Up to date, various routing algorithms are proposed, in both wired and wireless network. Generally, they can be classified as follows.

2.1.1 Distributed vs. Centralized Routing

Distributed routing system is often necessary in relatively large networks, in which the routing decisions are allocated to the wireless routers. In distributed routing, the routers exchange the information with each other, routing tables can be established based on these information. On the other hand, in centralized routing, it is the central processor that takes charge of all the routing in the network. With powerful cache implemented, the central processor designs the routing table for every station within its range, based on the link information it gathers, then it distributes the routing tables to wireless routers. Thus, centralized routing is suitable in networks with central administration system.

2.1.2 Source-Based vs. Hop-by-Hop routing

In source-based routing, the entire path from the source to the destination is decided at the sender station, and the path is implemented in the packet header, which leads the packet to the destination. Therefore, the source needs to have knowledge of the whole network topology, in order to pre-decide the routing correctly. However, this algorithm will make the packet disappear forever, if any link on the path goes down.

Hop-by-hop routing is more flexible on routing decisions, it is like a choose-as-you-go mode, and each intermediate router is able to choose the next hop, given only the neighbor's information, instead of the entire network. In the packet header, only the address of final destination is stored, instead of the whole path. Thus, the routing overhead is reduced in hop-by-hop mode, which is fundamental in the network with frequent topology updates.

2.1.3 Stochastic vs. Deterministic Routing

In stochastic routing, routers keep more than one choice for the next hop if applicable, on the way to the destination. Based on probability parameter or just random rate, the router will pick up a path to forward the packets. Consequently, the routing decision is made by the designed probabilistic model, and the destination may receive the data out of order. Deterministic routing avoids such problems, by forwarding packets to the destination along fixed paths.

2.1.4 Single-Path vs. Multi-Path Routing

Currently, most routing protocols are single-path, in which only one route is chosen between the source and the destination. Once a link along the path is broken, the whole path is not functioning any more. By contraries, multi-path routing maintains more than one route to forward the data to the same destination, normally all the routes are sorted or classified in a certain order, by which the router chooses the next available path, if primary path goes down. Furthermore, some multi-path routing algorithms support simultaneous transmissions along various routes between the source and the destination, thus re-ordering the packets at the destinations is quite necessary.

2.1.5 Dynamic vs. Static Routing

Dynamic routing often refers to routing algorithms that decide the next hop depending on current network situation or link state; it is also called state-dependent routing, which is very common in dynamic networks with frequent changes. This method can respond to network changes very fast and effectively, for instance, if the quality of a link on a path degrades, the algorithm may consider taking another node as the next hop. As the networks become various and complicated, dynamic routing shows its advantages on adjustability and scalability. On the other hand, some networks adopt static routing instead. Static routing does not take into account the network situation or link state, during network operation. So it contributes less overhead than dynamic routing protocols, as it does not need to send extra messages to collect the network status, however, this type of routing is only sufficient in networks with few or no changes.

As discussed above, routing in distributed systems can be explained as follows. Let the oriented topology graph G (V, E) represent an interconnected network, which has a set of vertex V and a set of edges E. In this weighted graph, each node in the set V represents a queuing or forwarding unit and each edge from E is a transmission system. The main task of an optimal routing algorithm is to direct data flow from source to destination nodes, maximizing network performance with certain routing protocol. Novel routing approaches are required to efficiently manage distributed multimedia services, mobile users and networks, heterogeneous inter-networking, service quality, point-to-multipoint communications, etc [6].

2.2. Related Works

In last decades, routing in wireless networks has become an attractive field, many researchers propose their characteristic protocols. Based on how path exploration is initiated, these algorithms can be classified into three categories: Global/Proactive, On-demand/Reactive and Hybrid method. Briefly speaking, Proactive protocols pave the way before the connection comes, and the route update process is performed periodically. Reactive routing only looks for certain path upon request of a new connection, which utilizes the resource efficiently. Recently, hybrid routing has shown its advantages under certain environment, it has the feature of both proactive and reactive routing, with the goal of maximizing scalability and efficiency. For instance, the reactive path setup and proactive path maintenance are combined together to achieve hybrid routing.

Destination-Sequenced Distance Vector (DSDV) routing [7] is a classical proactive algorithm, which guarantees loop free routes. It provides a single path to a destination, using the distance vector shortest path routing algorithm. However, DSDV introduces large amounts of overhead to the network due to the requirement of the periodic update messages, and the overhead grows according to network size. The protocol therefore will not scale in a large network. Global state routing (GSR) [8] is based on the traditional Link State algorithm, which has less overhead by restricting the update messages between intermediate nodes only. In GSR, every host keeps a link state table with its neighbors, based on received information from neighboring hosts, and periodically exchanges its link state information with neighbors only. However, the control messages are still relatively large, and they keep growing as the network scale. Thus, precious bandwidth is consumed by these update messages. Also inspired by traditional link-state algorithm, Optimized link state routing (OLSR) [9] is a point-to-point enhanced routing protocol. It minimizes the size of control messages and the number of rebroadcasting nodes by using Multi Point Replaying (MPR) strategy. During topology update, each node in the network selects a set of neighbors to retransmit its packets. This set of nodes is called the multipoint relays of that node. Any node which is not in the set can only read and process each packet. To select the MPRs, each node periodically broadcasts a list of its one hop neighbors using Hello messages. Most global routing protocols do not scale very well, due to their high overhead.

Reactive protocols, on the other hand, set up the route when it is needed, using their route discovery process. Dynamic Source Routing (DSR) [10] protocol requires each packet to carry the full address of a path, from source to the destination. This means that the protocol will not be very efficient in large networks, because routing overhead carried in the packet will continue to increase as the network diameter grows, which may consume most of the bandwidth. Ad hoc On-Demand Distance Vector (AODV) [11] routing protocol is based on DSDV and DSR algorithm. It uses the periodic beaconing and sequence numbering procedure of DSDV and a similar route discovery procedure as in DSR. The most distinguishing difference between DSR and AODV is that in DSR each packet carries full routing information, whereas in AODV the packets carry the destination address. This means that AODV has potentially less routing overhead than DSR. However, in AODV, node may experience large delays, and link failure may initiate another route discovery, which introduces extra delays and consumes more bandwidth as the mobility or size of the network increases.

Light-weight mobile routing (LMR) [12] protocol is another on-demand routing protocol, which uses a flooding technique to determine its routes. The nodes in LMR maintain multiple routes to each required destination. This increases the reliability of the protocol by allowing nodes to select the next available route to a particular destination without initiating a route discovery procedure. Another advantage of this protocol is that each node only maintains routing information to their neighbors. This avoids extra delays and storage overheads associated with maintaining complete routes. However, LMR may produce temporary invalid routes, which introduces extra delays in determining a correct loop.

Hybrid routing protocol is a new generation, which is both proactive and reactive in nature. This type of protocols always form a certain key backbone in the network, which can periodically update topology information of only most important areas, to reduce the routing overheads. Most hybrid protocols are zone-based, which partitions the network into a number of zones, while the other hybrid protocols group nodes together into trees or clusters, which differ from zone by assigning a control station in each cluster.

Zone routing protocol (ZRP) [13] is a typical zone hybrid protocol, the network has some routing zones, within which routes are maintained and immediately available for inside nodes. If the nodes lie outside the routing zone, routes are determined reactively, adopting any on-demand routing protocol. The advantage of ZRP is that the amount of routing overheads is largely reduced, while the delay in certain areas is minimized. However, ZRP needs a careful design with the number of routing zone, which is a tradeoff between scalability and efficiency.

Distributed Spanning Trees based routing protocol (DST) [14] represents the other group, which divides the network into a number of trees. In DST, each tree has two types of nodes: route node and internal node. The tree root controls the structure and operation of the tree, and the rest of the nodes within each tree are the regular nodes. DST utilizes spanning trees in the regions where the topology is stable, and resorts to an intelligent flooding-like approach in highly dynamic regions of the network. Then routing is performed using the spanning trees based on a hold-and-forward or shuttling method. They believe that as connectivity increases, and the network becomes more stable, it might be useful to buffer and route packets when the network connectivity is increased over time. Thus, in DST, control packets are sent to all the neighbors and adjoining bridges in the spanning tree, where each packet is held for a period of time called holding time. When a control message reaches down to a leaf node, it is sent up the tree until it reaches a certain height referred to as the shuttling level. After the shuttling level is reached, the control packet can be sent down the tree or to the adjoining bridges. The main disadvantage of the DST is that it relies on a root node to operate the tree, which may suffer from failure. Furthermore, the holding time referred may cause extra delays.

2.3. Maximum Routing Metric Design

To design a good routing algorithm, routing metric is quite important, which can be affected by lots of factors, possible influencing factors and the relationship are shown in Fig. 2-1. As seen from the figure, environmental factors are unidirectional to the network, which means they are not affected by the network, such as the placement and mobility of the nodes, and the technical properties or external interference. On the other hand, network factors are affected by the traffic within a network, like traffic congestion, internal interference, the consumed energy and also the topology of the network.

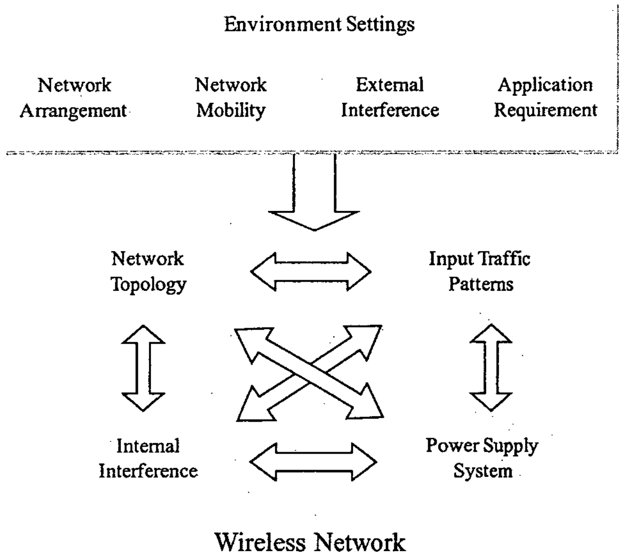


Fig. 2-1 Influence between environment and routing metrics

Besides the classic number of hops, there are many other types of routing metrics, which can be simply categorized as following: performance-based, radio-related, topology-based, mobility-based, geographic-related, and some other energy-related metrics.

Performance-based: If we desire better throughput, delay, jitter, or other quality issues, it is common that the related parameter is taken as a partial metric in the algorithm. Thus, metrics like delay, throughput, and packet loss ratio are often used in wireless routing algorithms.

Radio-related: Radio or signal is the base of wireless network, moreover, signal strength is considered as an indication for the link quality and the distance between two nodes. Thus, radio property is a popular metric.

Topology-based: This category includes the metrics produced from the network topology. In both ad hoc and mesh networking, hop count is the most commonly used routing metric in existing routing protocols such as DSR, DSDV, and AODV. This metric directly shows the effect of path length on the performance, and obtaining loop-free paths with minimum hop count is easy. However, it does not consider any quality issue of the path, so high-quality routes may not be found. Another interesting metric in this category is "Number of neighbors", which is defined by how many other nodes can be reached from a node at a given moment, it may indicate the surrounding condition for routing. However, if one node can reach the other, the other direction may not be the same, in another way, this metric is unidirectional.

Recently, Expected Transmission Count (ETX) was proposed in [15], which is defined as the expected number of transmissions a node requires transmitting a packet to a neighbor successfully. It is calculated using forward and backward delivery probability of probing packets. In ETX, each node broadcasts link probes of a fixed size, periodically at τ, and τ is jittered by up to 10% per probe to eliminate synchronization. All ETX probes are broadcast, so they are not acknowledged or retransmitted. Every node remembers the probes it receives during the last w seconds, allowing it to calculate the delivery ratio from the sender at any time *t* as:

$$d(t) = \frac{Count(t - w, t)}{w/t} \qquad (2\text{-}1)$$

where *Count (t-w, t)* is the number of probes received during the window w, and w/t is the number of probes that should have been received. Each node then maintains a table with the ETX to all the other nodes in the network based on its knowledge of the network. Later on, given different settings, more researchers became interested in this concept of ETX. Inspired by ETX, new metrics are proposed, like ENT [16], ETT [17] and WCETT [17], all of which have their own specific network environments and assumptions.

Comparison among these typical routing metrics is given in Table 2-1, in terms of load-balance, quality-aware, consideration of packet size, data rate, stability and measurability respectively. As analyzed, all these metrics have their own constraints and assumptions.

Table 2-1 Routing Metric Comparison

Metrics	Load-Balance	Quality-Aware	Packet Size	Data Rate	Stability	Measurability
Hop Count	No	No	No	No	Medium	Easy
Queue Length	Yes	Yes	No	Yes	Medium	Medium
ETX	No	Yes	No	No	Medium	Easy
ENT	No	Yes	Yes	Yes	High	Easy
ETT	No	Yes	Yes	Yes	Low	Difficult
WCETT	No	Yes	Yes	Yes	Low	Difficult

2.4. Load Balancing

Another important issue in designing a routing protocol for WMNs is load balancing, which is defined as evenly spreading the traffic load all over the network. Such a scheme is beneficial in the following aspects such as:

- Easing the traffic load on individual hosts and minimizing response time,
- Providing more reliability for the incoming connections,
- Ameliorating the wireless networks' security,
- Effectively utilizing precious network resource and optimizing throughput,
- Avoiding congestions at certain overloaded hosts.

As described, WMNs are the networks with many mobile clients roaming around, desiring various services from the network. In such an environment, it is quite possible and necessary for one mobile host to seek the aid of others in transmitting a packet to its destination, due to the limited propagation range of wireless radio. Therefore, the performance of the overall network may depend on the positions of the wireless stations and how a packet is routed to reach the destinations. Choice of the next hop is quite important in spreading the traffic fairly over the network, thus, a bad load-balancing scheme in routing may be a disappointment to the designer. Moreover, only load-balancing is not enough for optimal routing, real-time network status updates should be considered together, avoiding stale routing information.

Currently, load-balancing is a popular area in wireless telecommunication. To our best knowledge of the literature, ad hoc networks already have several mature load-balancing algorithms in stock. For WMNs, designing suitable load metrics is important to achieve load-balancing. In this thesis, we adopt a simple but useful metric, Transmission Queue Length (TQL), which is the available space in the transmission queue of the neighboring node. It is easy to measure and can reflect the traffic load on the next-hop node. In our proposed load-aware routing protocol as described in Chapter 4, the TQL information is exchanged between neighbors within communication range, using periodical “Ant” messages.

2.5. Estimated Bandwidth Methods

In addition to the load estimation, performance measurement is another way to evaluate the network. The most popular parameter of performance is bandwidth. Accurate network bandwidth measurement can make tremendous benefits, but accuracy is quite difficult to achieve due to some unexpected changes in the network. Consequently, using throughput as an estimation of bandwidth gains most popularity. However, throughput is defined as the amount of data the transport protocol can transfer per unit of time, so other parameters like packet drop rate may affect the throughput significantly, while not influencing bandwidth.

Also, quite a number of researchers prefer to use Pathchar [18], which sends packets of changing sizes and measures their round trip time to get estimation. It then combines the round trip times with the packet sizes to calculate bandwidth. Pathchar gives out pretty accurate measurement of bandwidth, which makes it a useful tool. However, Pathchar is slow and consumes much precious network bandwidth.

Other family of bandwidth measurement is called Packet Pair algorithm [19]. It relies on the fact that if two packets are queued next to each other at the bottleneck link, which is the link with minimum bandwidth, then they will exit the link t seconds apart:

$$t = \frac{S_2}{B_{bnl}} \qquad (2\text{-}2)$$

where S_2 is the size of the second packet and B_{bnl} is the bottleneck bandwidth, which is defined as the bandwidth of the bottleneck link on the path. Assuming the bottleneck separation is constant, the two packets will arrive at the receiver spaced t seconds apart. Since packet size S_2 is known, bottleneck bandwidth can be calculated:

$$B_{bnl} = \frac{S_2}{t} \qquad (2\text{-}3)$$

Based on this idea, some variants of basic Packet Pair algorithm came up with new features; they are Sender Based Packet Pair (SBPP), Receiver Based Packet Pair (RBPP), and Receiver Only Packet Pair (ROPP). As it can be seen from their names, t is measured at the sender, based on variance between arriving times of two ACK packets from the receiver, for SBPP; RBPP measures the time gap at the receiver, between arriving times of two packets; ROPP only takes timing measurements from the receiver and is therefore easier to deploy than RBPP. However, without timing information from the sender, ROPP cannot filter out time compressed packets or reordered packets, as SBPP and RBPP can [20].

2.6. Ant Colony Optimization Algorithm

Beginning in early 1990s, Ant Colony Optimization (ACO) [21] was first described as a new novel meta-heuristic for discrete optimization problems, inspired by the foraging behavior of real ants. Upon a task to search for food, nature ants initially explore the area surrounding their nest in a random manner. As soon as an ant finds a food source, it brings some samples back to the nest, with its own "evaluation" on the food. On the way back, the ant deposits certain chemical substance called "pheromone" on the road. The quantity of pheromone left depends on the quantity and goodness of the food, this behavior guides the other ants to the food source, also, this kind of communication between the ants help them to find shortest paths between their nest and food sources.

In the field of ACO routing, intelligent ants are transformed into a part of optimized routing techniques. The relevant ACO routing algorithms are actually stochastic search procedures, and the central component is the concept of "pheromone", which is used to probabilistically route the ants and data. The main schemes in ACO routing are introduced as follows [22]:

An ant is defined to be a simple computational agent, which iteratively constructs a solution for the problem to solve. For routing issue, every host the ant arrives is seen as a state, the ant moves from state i to feasible state j, according to a probability distribution.

In basic ACO routing, probabilities for choosing the next hop can be computed as follows [22]:

$$P\left(c_i^j \middle| \varsigma^p\right) = \frac{\left[\tau_i^j\right]^\alpha \cdot \left[\eta\left(c_i^j\right)\right]^\beta}{\sum_{c_k^l \in S(\varsigma^p)} \left[\tau_k^l\right]^\alpha \cdot \left[\eta\left(c_k^l\right)\right]^\beta}, \quad \vee\, c_i^j \in S(\varsigma^p) \tag{2-4}$$

where $S(\varsigma^p)$ is the host set with a partial host ς^p, the possible routing choice $c_i^j \in S(\varsigma^p)$ is done probabilistically with respect to the pheromone model, the probability for c_i^j is proportional to $[\tau_i^j]^\alpha \cdot [\eta(c_i^j)]^\beta$, where η is a customized function assigning to each c_i^j, and τ_i^j is the relevant pheromone value, $\alpha>0$ and $\beta>0$, determine the relative contribution of pheromone value and heuristic information.

Pheromone trail update: pheromone trail is defined as an end-to-end path between source and destination, with pheromone values on all links. It is updated at each iteration, increasing the pheromone level of those selected high-quality hosts, while decreasing all others [23]. Traditional ACO algorithm updates the trail as follows:

$$\tau_i^j(t) = \rho.\tau_i^j(t-1) + \Delta\tau_i^j \qquad (2\text{-}5)$$

where ρ ϵ (0, 1) is a user-defined parameter called evaporation rate, which is used to uniformly decrease all the pheromone values, to avoid extremely rapid convergence of the algorithm toward a sub-optimal situation.

Routing in ACO is achieved by transmitting ants rather than routing tables or by flooding Link-State-Advertisements (LSAs), thus, even though the size of an ant may vary in different systems or implementations, the size of ants is relatively small, in the order of 6 bytes [24]. However, Bonabeau et al. [25] have pointed out that the success of ants in collectively locating the shortest path is only statistical. If by chance, many of the ants initially choose a non-optimal, other ants are more likely to select leading to further reinforcement of the pheromone concentration along.

Since conventional ACO was not proposed for modern wireless networks, later on, many enhanced routing protocols were proposed based on ACO algorithm. Accelerated Ants Routing [26] randomly launches ant-like agents going through the network, but no destination is specified for them. On their way, the agents update pheromone entries pointing to their source. In [27], the authors presented Ant-AODV, which is a hybrid algorithm combining ants with the basic AODV behavior. In Ant-AODV, a certain number of ants travel around the network in a random manner, recording the last n visited nodes. The node routing table is updated upon the ants visit. Ant-Colony-Based Routing algorithm [28] establishes multiple paths between source and destination at the start of a data session with ants. During the data session, data packets confirm the paths they travel. In [29], Gianni Di Caro and Marco Dorigo proposed AntNet, which is a distributed, mobile agents based Monte Carlo system. It learns of routing tables adaptively, in order to react efficiently to changes in communications networks. In AntNet, agents concurrently explore the network and exchange collected information, the communication among the agents is indirect and asynchronous, mediated by the network itself, which exactly happens in the world of social insects and is called stigmergy.

Recently, many new routing protocols are proposed for various networks, based on ACO algorithm. AntHocNet is proposed in [30], it is designed for mobile ad-hoc networks. Although it achieves better performance than AODV, it only considers delay and number of hops as routing metric, thus the link quality is not fully estimated. Ants Intelligence Routing (AIR) [31] is proposed for WMN, it updates Round Trip Time (RTT) by exchanging "Hello" messages between neighbors, then the network local status is learned by the nodes. However, the routing choice is made mainly by certain value, which is randomly chosen, thus the estimation is not accurate. In [32], Node Neighbor Number Algorithm (NNNA) is proposed for Ad-hoc networks. NNNA uses the number of neighbors of a node to select the next hop, which may reduce the possibility of overload at certain nodes. However, the number of neighbors can not reflect the load situation accurately, because link quality does not depend on the number of neighbors at a node.

2.7. Summary

In this chapter the routing design and related issues are introduced, the routing algorithms are classified into various traditional categories generally. And then current popular routing protocols for wireless network are discussed as related works. Moreover, the basic concepts of our proposed routing algorithms are briefly introduced in this chapter, such as load-balancing, estimating bandwidth, and ACO algorithm. We use these concepts in designing our own routing protocols, which are presented in the next two chapters.

CHAPTER 3

Proposed ACO-based Routing Protocol: AIM

In this chapter, we propose our Ants-in-Mesh (AIM) routing protocol, which is based on ACO mechanism. It could be divided into several phases: on-demand path setup, data transmission, proactive path maintenance with link update and processing link failure. Next, simulation results are shown and analyzed, in terms of packet delivery ratio, end-to-end delay and overhead. We compare our work with AODV, AIM shows better performance than AODV, in both network mobility and scalability.

AIM is a single-radio multipath routing algorithm, which is mainly designed for wireless mesh networks. AIM embeds multifunctional ant agents on each wireless mesh router and the mesh client. These agents are utilized to schedule and launch different types of ants (reactive and proactive, forward and backward). End-to-end routes are not memorized at all times, AIM adopts the popular on-demand scheme for path setup. Certain numbers of reactive forward ants are scheduled by an ant agent on the source in order to find multiple paths to the destination when the data session starts. Receiving multiple forward ants, the destination will decide which ones could be used, and then launch backward ants to confirm the paths. After the route setup, data sessions begin to send out data packets to the destination along the confirmed paths. For the maintenance of route, AIM monitors the end-to-end paths by periodically sending out proactive forward ants. Since the topology may change randomly, the local link status within one hop are updated with "Hello" messages. When links fail, AIM will send the notification along the paths to warn the relevant preceding nodes.

In the following sections, we present the details of the AIM routing protocol.

3.1. Ants Routing Model

Let the oriented topology graph G (V, E) represent an interconnected network, which has a set of vertex V and a set of edges E. According to the ant algorithm, τ_{ij} (t) is defined to be the density of pheromone on link (i, j), then set $\Gamma\{ \tau_{ij} (t) \mid e_{ij}, e_{ij} \in E\}$ represents the residual pheromone on links E at any moment t. AIM routing is then carried out on the oriented graph G (V, E, Γ).

Pheromone table is updated in the following manner:

$$\tau_{ij} (t+1) = (1-\rho)\, \tau_{ij} (t) + \Delta\tau_{ij} (t) \qquad (3\text{-}1)$$

At moment t+1, pheromone value on link (i, j) equals to evaporated pheromone plus the increment of pheromone. Here, ρ is an evaporation coefficient of pheromone, 1-ρ indicates the residual of pheromone on this link, while $\Delta\tau_{ij}$ (t) means increment of pheromone. In AIM, per-hop pheromone can be updated by both backward ants and Hello messages, and the increment of pheromone can be obtained by the following equations:

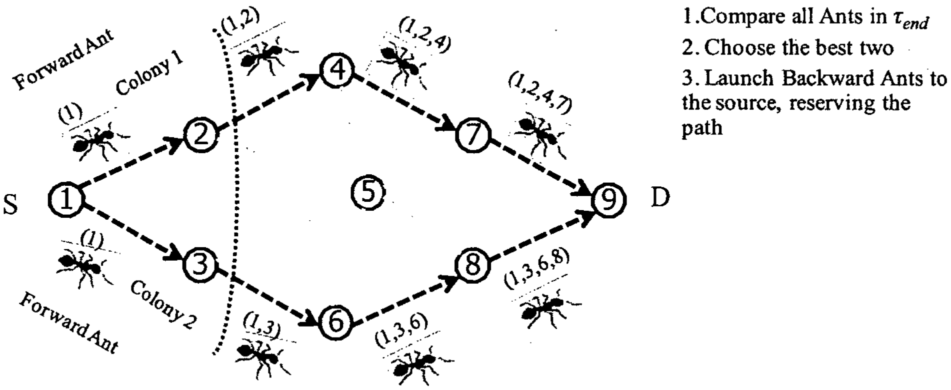
$$\Delta\tau_{ij} (t) = \alpha \cdot F\left(1- \frac{T_{delay}}{T_{thres}}\right) + \beta \cdot F\left(1- \frac{N}{N_{thres}}\right) \qquad (3\text{-}2)$$

$$F(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \qquad (3\text{-}3)$$

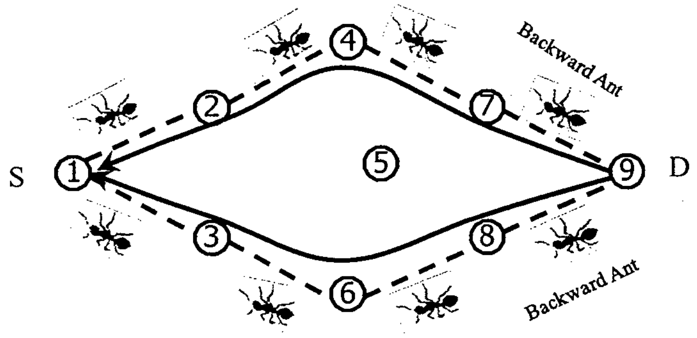
where T_{delay} means the per-hop round trip time (RTT), N is the node's connection number, which reflects how many connections this node has with the neighbors. If a link is occupied by a confirmed path, then we say there is a connection between the two nodes of the link. T_{thres} and N_{thres} are the relevant thresholds, which constrain the pheromone values. α and β are the coefficients (set to 1.5 and 1 in the simulation) of per-hop RTT and connection number respectively.

3.2. On-Demand Path Setup

At the first phase, when a source node *s* wants to communicate with the destination node *d*, but no routing information of *d* is available on *s*, *s* needs to search for paths connecting *s* and *d*. It broadcasts a reactive forward ant FA[s,d] to all the neighbors within radio range. Each FA[s,d] carries colony identification, generation number, source and destination address, timestamp and an empty ant stack which can be used to record the forward route. Generation k is defined as the kth hop from the source node, every first generation ant acts as the queen ant of their own colony, and each colony has its own identification in the ant for further utility. Thus, the number of neighbors of the source node is the number of the colonies. The final mission of all the colonies is to find multiple link disjoint paths to the destination.



(a) Forward Ant searching the paths



(b) Backward Ant confirming two paths

Fig. 3-1 On-Demand Path Setup in AIM

At each intermediate node, when the agent receives a FA[s,d], it will push the current node's ID into stack to record the route. Meanwhile, the agent checks its routing table to find the next hops with best pheromone values. As soon as it finds several best next hops to go, this generation ant will replicate limited number of next generation ants, and then these new born ants are sent out to respective next stations. With the above policy, ants can proliferate quickly over the network, following different paths to the destination. The path setup is shown is Fig. 3-1.

However, there is a possibility that the intermediate nodes may receive the same colony's ants, but with younger generation, this is called a routing loop. Under such circumstances, the agent just discards the ant. On the other hand, if the ant's travel time or number of hops exceeds the limit, it is also killed by the agent.

In order to prevent certain nodes from being overused, and to utilize the available network resources more efficiently, the destination must know the entire path of all available routes so that it can select the routes according to certain routing metrics [33]. Additionally, what we focus on is to achieve k best end-to-end multiple paths, thus, there is no route comparison or decision made by the intermediate stations. Only the destination can terminate the search process of forward ant and launch a backward ant to confirm the path. There is no need to store routing information at intermediate nodes in the proposed approach, in comparison to the AODV and DSR, where the intermediate nodes store the routing information to the destinations. The destination node may not know the information of the route, consequently it is difficult to establish maximally disjoint multiple routes in DSR and AODV protocols [33].

Upon arrival at the destination *d*, the visited nodes stack of this forward ant is unpacked by the destination agent, and also the route pheromone information is stored for comparison with other ants, if this ant is qualified (means this ant's pheromone is within certain range). At the same time, these ants are sorted by route quality in the pheromone table. End-to-end pheromone for each path is calculated by the equations below:

$$\tau_{end} = \alpha \cdot F\left(1 - \frac{T_{delay}}{T_{thres}}\right) + \beta \cdot F\left(1 - \frac{N_{\max}}{N_{thres}}\right) + \gamma \cdot F\left(1 - \frac{H}{H_{thres}}\right) \qquad (3\text{-}4)$$

$$F(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \qquad (3\text{-}5)$$

where T_{delay} means the end-to-end round trip time (RTT), N_{max} is the maximum node connection number on this route. At the destination, H is defined as the number of hops between the source and destination. T_{thres}, N_{thres} and H_{thres} are the relevant thresholds (set to 0.2s, 20, and 30 in the simulation), which constrain the pheromone values. And α, β and γ are the coefficients (set to 2, 1.3, and 1 in the simulation) of RTT, connection number and number of hops respectively.

Every destination has an increment timer and a counter for the forward ants coming from the source. The timer and the counter begin when the first qualified ant arrives. The counter augments by one whenever a qualified ant arrives. When the counter or the timer reaches the threshold, the destination input is terminated to any further forward ants from the same source; and those qualified forward ants are fully converted into backward ants, which travel back to the source confirming the path. However, if the backward ant cannot find next hop, due to node movement or link failure, it is deleted. On the way back to the source, the backward ants will update pheromone using equations (3-4) and (3-5), and reserve the bandwidth by updating connection number of involved nodes.

Generally, the number of the source's neighbors, N, decides how many end-to-end paths may be obtained through the ant exploration. In order to reduce overhead, we restrict the number of desired paths to L. In this way, this number (defined as colony number) will be either N or L, whichever is less. However, due to the network condition, the final confirmed paths may be less than colony number.

The structure of the source node is similar to that of destination, with a timer and a counter, they begin to count when the first backward ant arrives. All the backward ants are stored in terms of corresponding pheromone values and confirmed paths. When the source node receives enough backward ants or the timer reaches its limit, it starts to transmit the data. With rare possibility, if the source cannot receive any backward ant within certain time, forward ants will be launched again to search for route.

3.3. Data Transmission

After the success of path setup, the source node learns several good routes to the destination, with their own pheromone values. The waiting traffic data will be allocated on multiple disjoint paths according to the proportion of the pheromone for different paths. The higher pheromone a path has, the more data will be assigned on it, Fig. 3-2 shows how it works. As soon as the data session begins, there is no need for the data packets to decide which next hops to go, they are forwarded along the confirmed paths until they reach the destination.

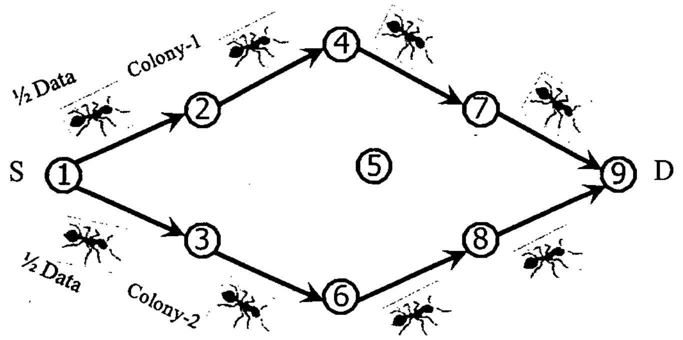


Fig. 3-2 Data Allocation in AIM

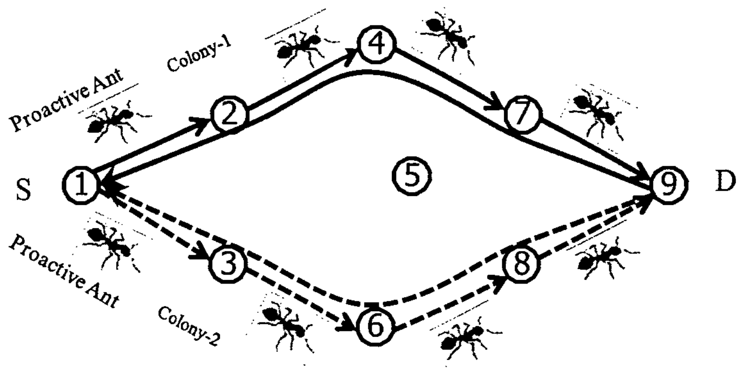
With this strategy, we do not need to calculate the pheromone or the probability for next hop each time the data arrives at an intermediate node. More efficiently, the data load is distributed over the network, with respect to estimated quality of the paths. However, owing to the variety of the network topology and utilization, the estimated quality may be stale, which can cause huge number of data loss. Thus, AIM enables path verification and update, and the details are discussed in the next section. With updated messages, when the confirmed path degrades or upgrades, the traffic data will be reallocated to still keep the same proportion. Consequently, the overloaded path will relieve from congestion, and the demanding path can get more traffic.

3.4. Proactive Path Maintenance and Link Detection

As discussed above, once the route is established, data will be transmitted between source and destination. However, because of the changes in network topology, we need an up-to-date view of the network. Therefore, periodical path maintenance and per hop link detection are required.

3.4.1 Proactive Path Maintenance

In AIM, proactive ants with a certain rate are used to verify the path. These ants are unicast like the data packets, and serve two main purposes. One is to confirm that the path is still valid; the other is to update end-to-end pheromone tables at both the source and the destination. Whenever a proactive ant arrives at the on-path node successfully, it gathers the pheromone information. Upon arrival at the destination node *d*, the proactive ant unloads all the pheromone information to refresh the pheromone table. After that, the destination agent will launch a backward ant as it did for reactive forward ants. This backward ant has only one task, which is to update the pheromone table at the source node. Hence, the source node can reallocate the traffic according to the new pheromone table. The proactive path maintenance phase is shown in Fig. 3-3.



1.Refresh the End-to-End Pheromone table
2.Update the Pheromone table at each node

Fig. 3-3 Proactive Path Maintenance in AIM

3.4.2 Link Detection

Both mesh routers and mesh clients need to keep the local information of their neighbors, including per-hop RTT, available bandwidth, and also the calculated pheromone value. In order to make the network more real-time and precise, Hello messages are used to detect links between neighbors and update routing information.

The main structure of Hello message contains the address of the sender (sender_addr), the message sent time (send_time), and the available bandwidth (represented as the confirmed connection number of sender). These Hello messages are broadcast every T_{hello} seconds (e.g., T_{hello} = 1 sec) by the nodes. Receiving a Hello message, the node agent calculates the RTT by subtracting the timestamp from current time. It further checks if the neighbor information for this sender is available. If available, it updates the relevant values in its neighbor table. Otherwise, it creates an entry in the neighbor table for this new neighbor.

Accordingly, the node responds by sending back a Hello message with its own information to let the sender update its neighbor table. After this first communication, all involved nodes expect to receive a Hello from the neighbors every T_{hello} seconds. Link detection is shown in Fig. 3-4.

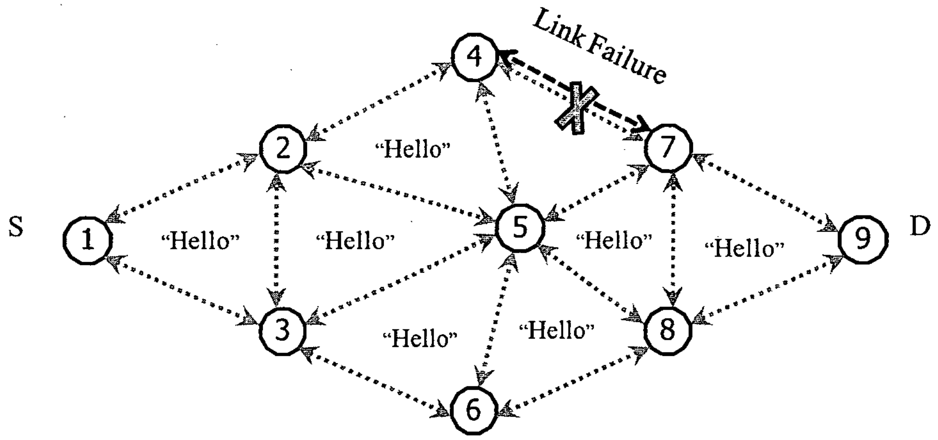


Fig. 3-4 Link Detection in AIM

3.5. Link Failure

In a wireless mesh network, AIM updates the routing view between immediate neighbors by periodically exchanging Hello messages, as shown in Fig. 3-5. Thus link failures can be detected as quickly as possible. The existence of a neighbor station can be confirmed when a greeting from this neighbor is received, or the success of a packet transmission. If there is no response or action from a neighbor for a defined period of time, which is T_{hello} × Loss Threshold (set to 2 in our simulation), the neighbor is simply removed from the neighbor table and the routing table. For another instance, if the scheduled controlling or data packet transmission fails, the two ends of this failed link will send out notification, which is destined to both source and destination. Finally, when the source and destination receive this notification message, they will delete this path and reallocate the traffic, based on original pheromone proportional values. As soon as the number of missing routes reaches a threshold (set to 2 in the simulation), the source will start the path exploration without affecting the current available routes. Therefore, AIM reacts efficiently to link failure while keeping the overhead low.

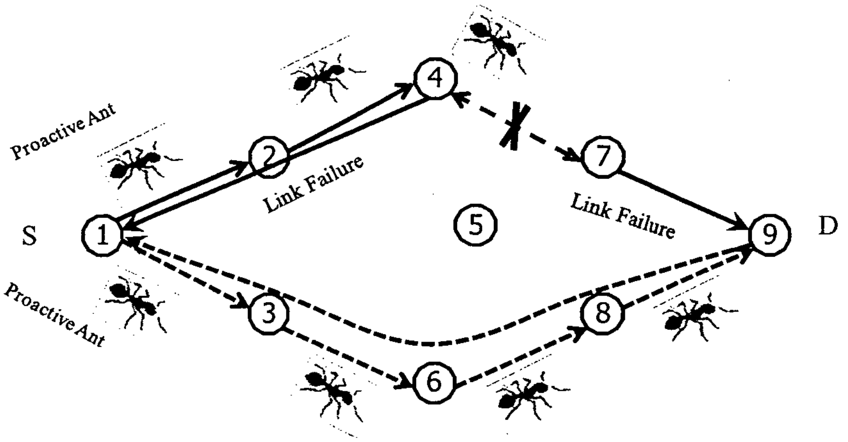


Fig. 3-5 Deal with Link Failure in AIM

3.6. Simulation Experiments

In order to validate our AIM protocol in wireless mesh networks, interlaced simulation experiments with combined environments are required. QualNet [34] is chosen as our simulator due to its accurate models for all the layers and the compatibility. AIM performance is compared with that of AODV (without local repair), which is one of the most popular routing protocols in current mesh networks.

3.6.1 Simulation Environment

Owing to the variety of simulation requirements, we conducted a wide range of experiments based on two different basic environments. The topology for the first experiment is shown in Fig. 3-6, which is popularly adopted for simulation.

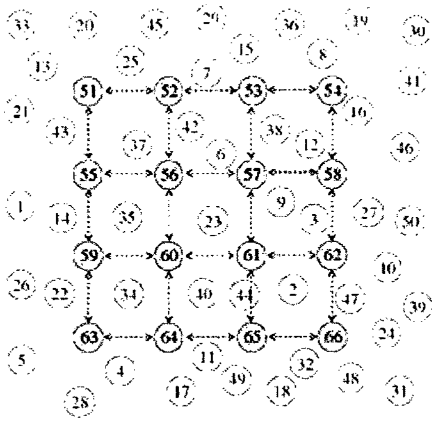


Fig. 3-6 Structure of mesh network for AIM simulation

The simulation parameters for the first experiment are defined in Table 3-1. For each scenario, we simulated the network 20 times with different random seeds, which can result in enough randomicity and accuracy. Finally, we presented the graphs with an average of 20 experiments.

Table 3-1 First Experiment Parameter for AIM

Examined Protocols	AODV AIM
Simulation Time	900 seconds
Network Size	1500 * 1500 M^2
Number of Mesh Routers	16
Number of Mesh Clients	50
Movement Pattern	Random Waypoint
Propagation Model	Two-Ray
Radio Range	298.56 M
Data Rate	2.0 Mbps
MAC Layer Protocol	802.11 b DCF
Traffic Type	CBR (UDP)
Maximum Connections	20
Data Session Start Time	Random Number in (0,180)
Data Session End Time	End of Simulation
Transmission Rate	64 Bytes/Second
Node Speed	0, 10,20,30,40,50 M/S

For the second setting, we evaluate the protocol scalability by varying the network size. In our case, the concrete parameters used for the network topology and the size of the area are given in Table 3-2. The data traffic consists of 20 CBR sources sending one 512-byte packet per second. The nodes move according to random waypoint model [35], with a minimum speed of 0 m/s, a maximum speed of 10 m/s, and a pause time of 30 seconds. Each experiment is run for 400 seconds.

Table 3-2 Network Environments for AIM

Area Size(m^2)	Clients	Routers	Total Nodes
1500 × 1500	50	16	66
1900 × 1900	75	25	100
2100 × 2100	100	36	136
2600 ×2600	150	49	199
3000 × 3000	200	64	264

3.6.2 Simulation Results

3.6.2.1 Effect of Speed on Mobility

Fig. 3-7 shows that, as the speed of nodes goes up, the PDR of both AODV and AIM degraded. The PDR of AODV decreased from 98% to 75%. On the other hand, AIM has less reduction in PDR, from 100% to 82%.

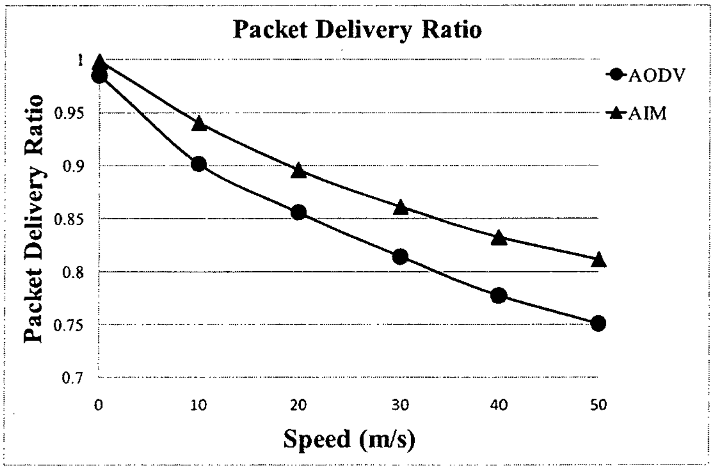


Fig. 3-7 Packet Delivery Ratio of AODV and AIM at different speeds

The total end-to-end delay in the network is compared in Fig. 3-8. As shown, AIM outperformed AODV at any speed.

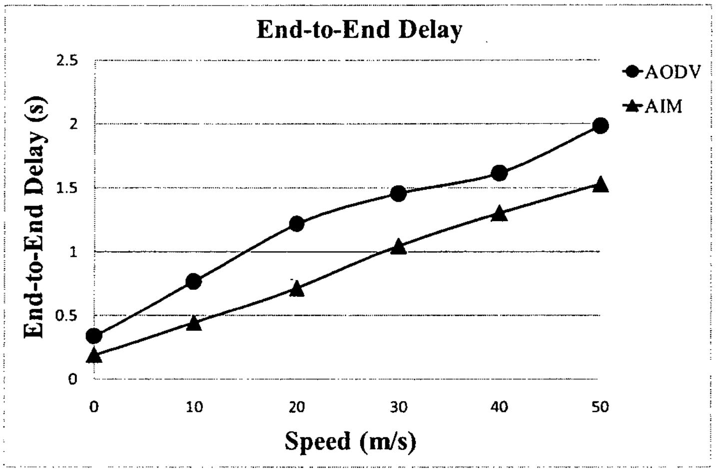


Fig. 3-8 Total End-to-End Delay of AODV and AIM at different speeds

However, better achievement comes at a cost. As shown in Fig. 3-9, when the clients become more mobile, the overhead of AODV increased with a small grade. In contrast, AIM experienced a steep climb on overhead.

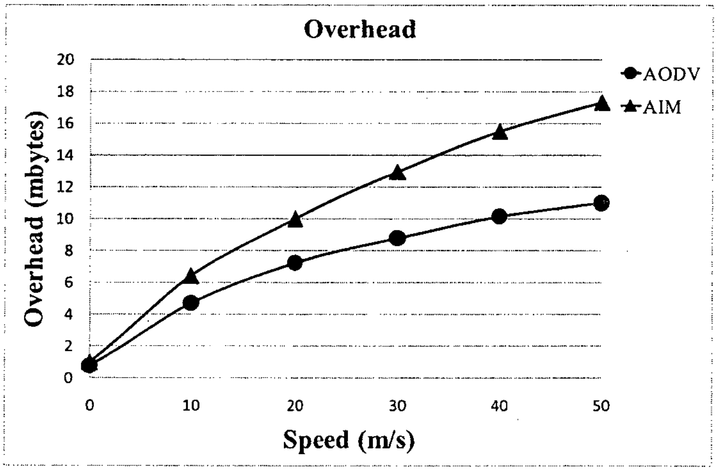


Fig. 3-9 Overhead of AODV and AIM at different speeds

3.6.2.2 *Effect of Pause Time on Mobility*

Figs. 3-10 and 3-11 show the results of network performance at different pause time of mobile clients. As seen from the figures, packet delivery ratio increases as the clients pause longer, while the end-to-end delay goes down.

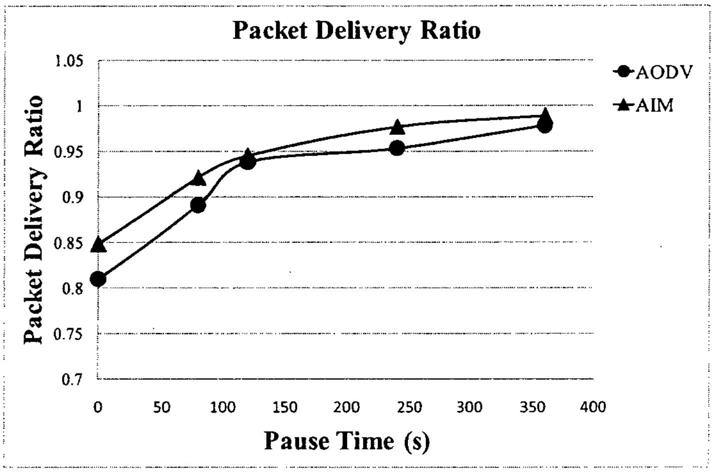


Fig. 3-10 Packet Delivery Ratio of AODV and AIM with different pause times

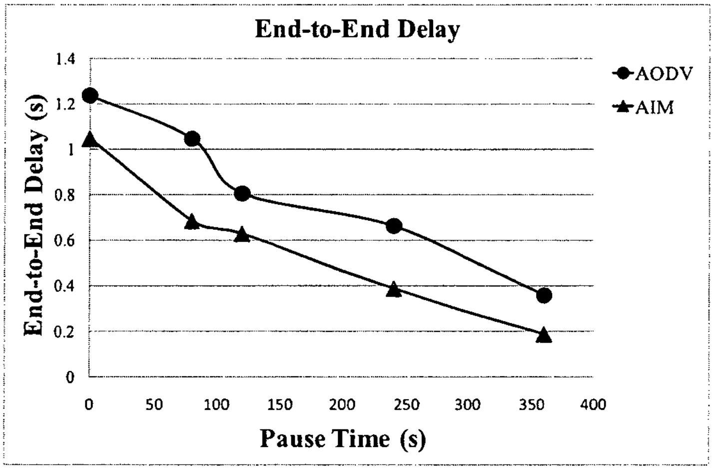


Fig. 3-11 Total End-to-End Delay of AODV and AIM with different pause times

3.6.2.3 Network Size

Fig. 3-12 presents the curves for packet delivery ratio as a function of network size. AIM achieved better performance than AODV in all network sizes. With respect to total end-to-end delay in Fig. 3-13, after the network size exceeds 136 nodes, AODV suffers from very high delay, while AIM experiences a relatively gentle increase in delay.

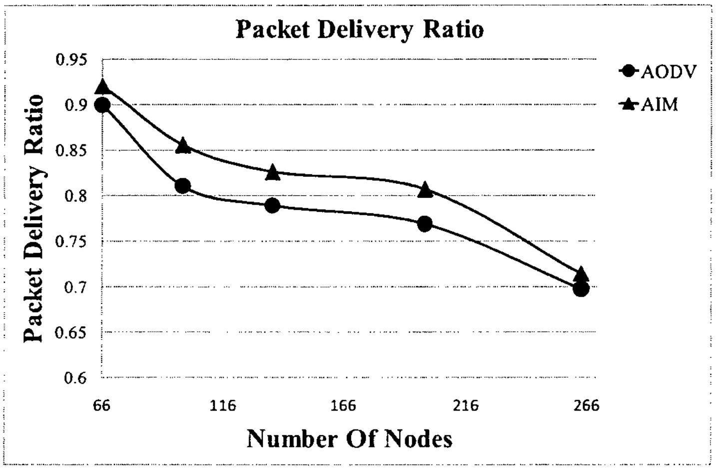


Fig. 3-12 Packet Delivery Ratio of AODV and AIM at various network sizes

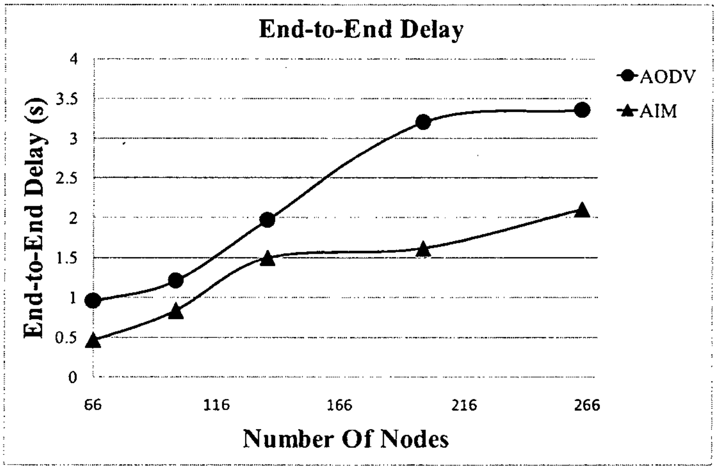


Fig. 3-13 Total End-to-End Delay of AODV and AIM at various network sizes

3.7. Summary

In this chapter, we present the details of our AIM routing protocol. AIM is a hybrid routing protocol, with reactive path setup and proactive path maintenance. Thus, network resource is utilized efficiently and AIM can react fast to topology changes. Using "Hello" messages, mesh stations are aware of their neighbors and the ongoing traffic between neighbors, this scheme updates network status only between neighbors. Moreover, connections between the station and its neighbors can reflect how much load is carried at the station. We use multi-path routing mechanism in AIM, a source-destination pair can have multiple routes to transmit the traffic, according to specific proportion value for each path. Consequently, data loss due to link failure is minimized and network performance is improved.

We simulate AODV and AIM routing protocols in Qualnet, the simulation results show AIM performs better than AODV in scalable mobile network, in terms of packet delivery ratio and end-to-end delay. However, AIM witnesses more overhead than AODV, because AIM utilizes various Ants to maintain the network status, collect and exchange information.

CHAPTER 4

Proposed ACO-based Routing Protocol: LAIM

This chapter introduces our proposed Load-aware-Ant-In-Mesh (LAIM) routing protocol, which mainly has three phases in it: set up the path upon demand, carry on the data transmission and path maintenance. After explaining the algorithm, we show and analyze the simulation result in this chapter.

Load-aware-Ant-In-Mesh (LAIM) is a single-radio multipath-enabled routing algorithm, also based on ACO algorithm. It could be seen as an enhanced version of AIM with more accurate update and more powerful functions.

In AIM, "Hello" message is used to exchange information like delay and connection number of a node, between neighbors. However, these two parameters are not accurate or good enough to reflect the network status. In LAIM, we propose new "Ant" message, which collects more information, such as bandwidth, delay, loss ratio and TQL, these parameters are then used to compare the goodness of a path. What is more, although AIM may balance the load with multiple paths, it still suffers from congestion, because path maintenance for multiple routes requires extra control messages, which occupy much network bandwidth resource, and the traffic load is not allocated efficiently in the network. Therefore, we choose best two paths after path exploration in LAIM, to keep overhead low. And more accurate bandwidth estimation method is adopted in LAIM, which also avoids overloading the mesh stations by using transmission queue length in metric design.

Since the topology may change randomly, the local link status within one hop is updated with "Ant" messages, which will also update the neighbors' quality information (like estimated bandwidth, delay, loss ratio, and TQL) in neighbor table. When links fail, LAIM will send the notification along the paths to warn the relevant preceding nodes.

4.1. LAIM Routing Model

Let the oriented topology graph $G\ (V,\ E)$ represent an interconnected network, which has a set of vertex V and a set of edges E. In LAIM, traditional ACO algorithms' pheromone table is modified and embedded to fit WMNs. Accordingly, τ_{ij} is defined to be the density of pheromone on links (i, j), then $\Gamma\ \{\tau_{ij} \mid e_{ij},\ e_{ij} \in E\}$ represents the current pheromone on links E, LAIM routing is then carried out on the oriented graph $G\ (V,\ E,\ \Gamma)$. τ_{ij} can be calculated and updated by the following manner:

$$\tau_{\mathrm{ij}} = (\frac{k_1}{Band_{ij}} + k_2 \cdot D_{ij} + k_3 \cdot ETT_{ij}) \,/\, TQL_{ij} \qquad (4\text{-}1)$$

where k_1, k_2, k_3 are the coefficients (set to 2, 1, 1 in the simulation) to balance different parameters, $Band_{ij}$ is the estimated bandwidth units between i and j, D_{ij} is the estimated communication delay units on link (i, j), ETT_{ij} is the Expected Transmission Time units, TQL_{ij} is a very important parameter called Transmission Queue Length, which is shown in Fig. 4-1. Periodically, pheromone value τ_{ij} on link (i, j) is updated by "Ant" message.

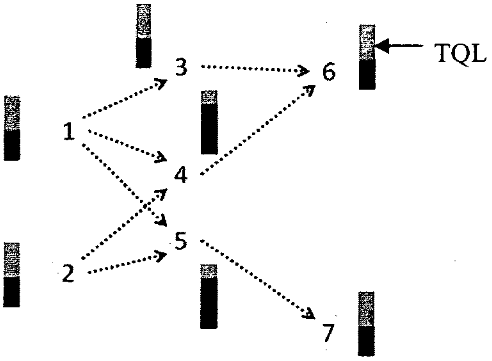


Fig. 4-1 Transmission Queue Length of each node

At the destinations, all the available paths will be filtered until two best paths are chosen, depending on the following formula:

$$\tau_{end} = \left(\frac{e_1}{Band_{bnl}} + e_2 \cdot D_{end} + e_3 \cdot ETT_{end}\right) / TQL_{bnn} \qquad (4\text{-}2)$$

where e_1, e_2, e_3 are the coefficients (set to 2, 1, 1 in the simulation) to balance different parameters, $Band_{bnl}$ is the estimated bottleneck bandwidth units, D_{end} is the end-to-end delay units on this path, ETT_{end} is accumulated Expected Transmission Time units, and TQL_{bnn} is the bottleneck Transmission Queue Length that has the minimum TQL along the path. In the worst case, the approximate values of $Band_{bnl}$, D_{end}, and ETT_{end} are 0.2, 10, and 10 respectively. Thus, we set (e_1, e_2, e_3) to (2, 1, 1), to keep all the parameters at the same level. From the source to the destination, forward ants collect various information along the path, then pour them into the destination's filter, to get the best two τ_{end}, as shown in Table 4-1.

Table 4-1 Destination Path Table

Available paths (In order of goodness)	τ_{end}
Path 1 (1, 2, 3...)	0.8
Path 2 (1, 4, 2...)	0.6
Path 3 (.........)	...
Path 4 (.........)	...

In [17], ETT of a link is defined as a "bandwidth-adjusted ETX". In other words, ETX is combined with the link bandwidth and packet size, to obtain the time spent in transmitting the packet. It is formalized as follows, let S denote the size of the packet and B the bandwidth of the link. Then:

$$ETT_{ij} = ETX_{ij} \cdot \frac{S}{B}, \quad ETT_{end} = \sum_{ij \in path} ETT_{ij} \qquad (4\text{-}3)$$

The ETX metric incorporates the effects of link loss ratio, asymmetry in the loss ratios between the two directions of each link, and interference among the successive links of a path, the ETX of a link is calculated using the forward and reverse delivery ratios of the link. The forward delivery ratio, d_f, is the measured probability that a data packet successfully arrives at the recipient; the reverse delivery ratio, d_r, is the probability that the ACK packet is successfully received. Then:

$$ETX_{ij} = \frac{1}{d_f \times d_r} \qquad (4\text{-}4)$$

In LAIM, nodes track the number of "Ant" messages received from each neighbor during a sliding time window w and include this information in their next "Ant". In this way, all the nodes can calculate d_f directly from the number of "Ant" they receive from a neighbor in the time window, and they can obtain d_r through the last "Ant" from a neighbor.

Generally, LAIM consists of two aspects in the routing metric: network condition and load situation, which are represented respectively by performance metrics including bandwidth, delay, and packet loss, and by load metric Transmission Queue Length, which can avoid overusing certain centralized nodes.

4.2. On-Demand Path Setup

LAIM is quite similar to AIM at the first phase of path setup, which is introduced in previous chapter. Additionally, the launched forward ants are required to gather necessary information along the path, like D_{end}, $Band_{bnl}$, TQL_{bnn}, to calculate ETT_{end} at the destination, as shown in Fig. 4-2.

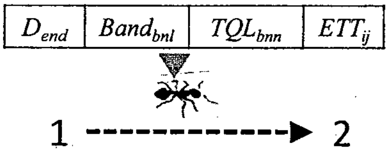
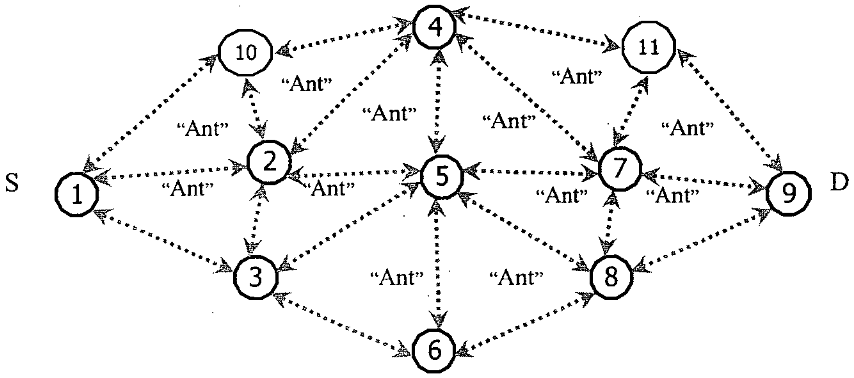


Fig. 4-2 Ant Collects Information

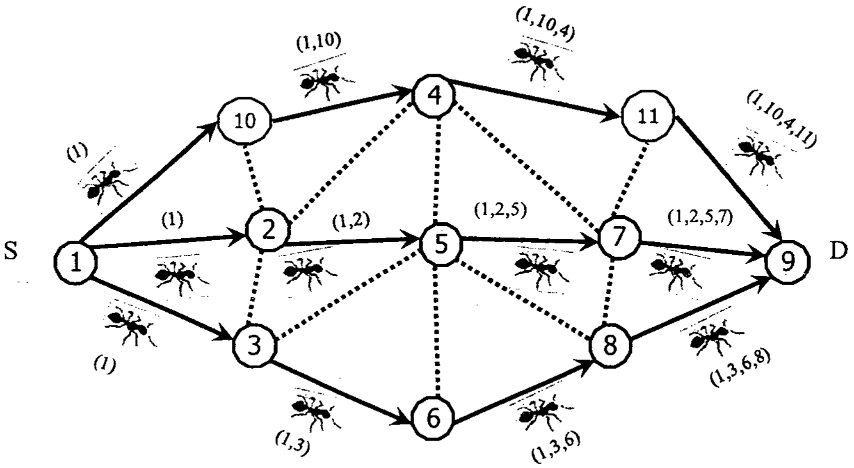
At each chosen intermediate node, when the agent receives a FA[s,d], it will push the current node's ID into stack to record the route. Meanwhile, the agent checks its neighbor table to find the next hop with best pheromone values, which includes estimated bandwidth, delay, loss ratio, and the TQL of the neighbor for load balancing. As soon as it finds it, this generation ant will give birth to a next generation ant, and then it is sent to next hop. In this way, the routing loop problem can be avoided. However, if the ant's travel time or number of hops exceeds the limit, it is killed by the agent.

In addition to TQL metric, in order to prevent certain nodes from being overused, and to utilize the available network resources more efficiently, the destination must know the entire path of all available routes so that it can select the routes according to certain routing metrics. Additionally, routing decision cannot be made by the intermediate stations, only the destination can terminate the search process of forward ant and launch a backward ant to confirm the path.

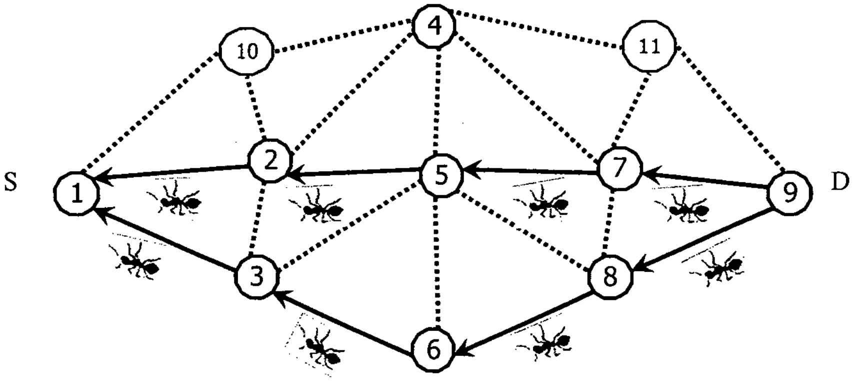
Upon arrival at the destination *d*, the visited nodes stack of this forward ant is unpacked by the destination agent, and also the end-to-end pheromone information is stored for comparison with other ants, if this ant is qualified. At the same time, these ants are sorted by route quality in the pheromone table. With this step, we can filter the routes by users' application requirement, like bandwidth, delay, or loss ratio. On-demand path setup phase is shown in Fig. 4-3.



(a) Exchange "Ant" messages between the neighbors



(b) Forward Ant searching the possible paths



(c) Backward Ant confirming the two chosen paths

Fig. 4-3 On-demand Path Setup in LAIM

Like AIM, every destination has an increment timer and a counter for the forward ants coming from the source. The timer and the counter begin when the first qualified ant arrives. The counter augments by one whenever a qualified ant arrives. When the counter or the timer reaches the threshold, the destination door is shut from the same source; and those qualified forward ants are fully converted into backward ants, which travel back to the source confirming the path. However, if the backward ant cannot find next hop, due to node movement or link failure, it is deleted. On the way back to the source, the backward ants will reserve the link by updating the neighbor table. Consequently, the following new connections can obtain updated routing information.

Normally in LAIM, the number of the source's neighbors, N, decides how many end-to-end paths may be obtained through the ant exploration. However, due to the network condition, the final qualified paths may be less than colony number, and in LAIM, we just choose two best paths, one is the primary, while the other is the backup.

4.3. Data Transmission and Backup Path

After the success of path setup, the waiting traffic data will be sent out along the confirmed paths to the destination. With this strategy, we do not need to calculate the pheromone or the probability for next hop each time the data arrives at an intermediate node, as in traditional ACO algorithm. In LAIM, two final paths are chosen, one for primary data transmission, while the other is used as backup, whenever there is failure on the primary one. Data transmission phase is shown in Fig. 4-4.

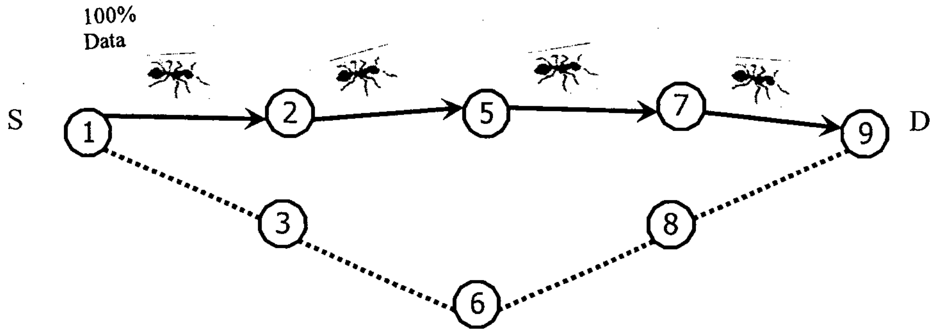


Fig. 4-4 Data transmission on the primary path in LAIM

4.4. Status Update and Link Detection

Due to the changes in wireless mesh networks, many transmissions may be cancelled or delayed. In order to reduce loss maximally, periodical update should be performed on all the links and the network status.

4.4.1 Proactive Path Maintenance

In LAIM, proactive ants are used to verify the paths, gathering necessary information on their way to the destination, as shown in Fig. 4-5. As soon as the proactive ants arrive and unload at the destination, the agent will launch a backward ant to inform the source of any changes, if s does not receive any news from the *d* in certain time, it assumes the path is down. If the primary path is down, s will just send data along the backup path, meanwhile, it will search for another good path to be primary or backup, depending on the goodness of the path. If the backup is down, the same search procedure will be performed.

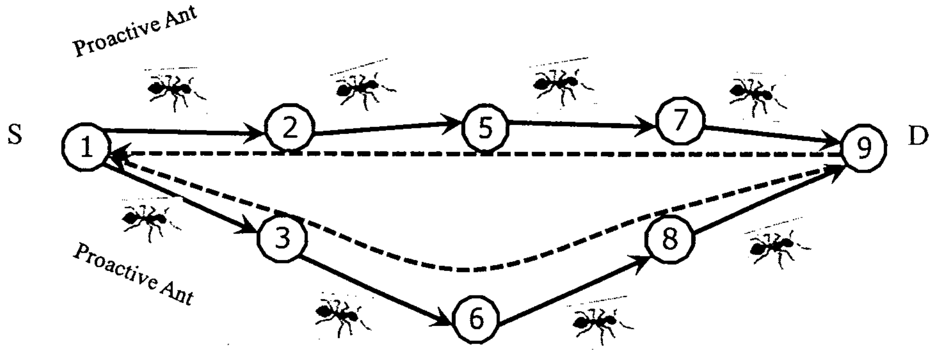


Fig. 4-5 Proactive Path Maintenance in LAIM

4.4.2 Link Detection

Both mesh routers and mesh clients need to keep the local information of their neighbors, including per-hop delay, available bandwidth, loss ratio and also the neighbors' TQL. In order to make the network more real-time and precise, "Ant" messages are used to detect links between neighbors and update routing information. The main structure of "Ant" message contains the address of the sender, the message sent time, and the local information. These "Ant" messages are broadcast every T_{ant} seconds by the nodes. They check if the neighbor entry for this sender exists. If available, it updates the relevant values in its neighbor table; otherwise, it creates an entry in the neighbor table for this new neighbor. After this first communication, all involved nodes expect to receive a "Ant" from the neighbors every T_{ant} seconds. In case of link failure, the same procedure is carried out in LAIM as in AIM. The link detection procedure is shown in Fig. 4-6.

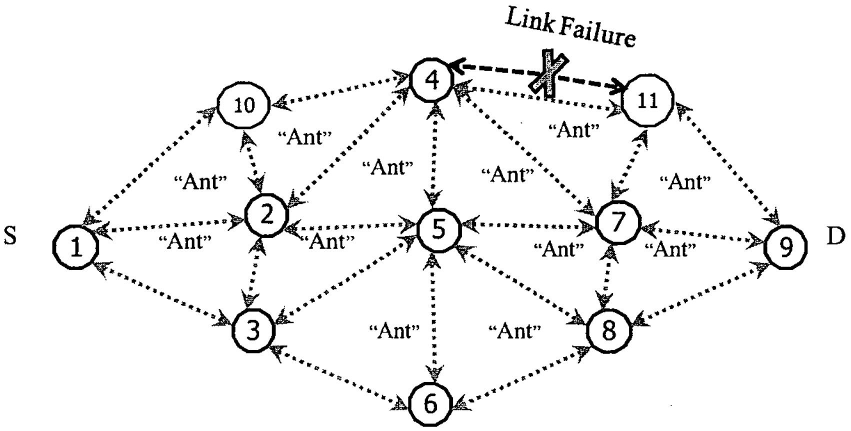


Fig. 4-6 Link Detection with Ant Message in LAIM

4.5. Simulation Experiments

In the simulation, LAIM performance is compared with that of AODV and AntNet. The three algorithms are compared with respect to several metrics: packet delivery ratio, total end-to-end delay and routing overhead ratio, which is calculated by dividing control bytes over data bytes.

4.5.1 Simulation Environment

Owing to the variety of simulation requirements, we conducted a wide range of experiments based on the two different basic environments.

The topology for the first experiment is shown in Fig. 4-7, which consists of 50 nodes, with 9 mesh routers and 41 mesh clients, the core of the WMN is the mesh routers, while all the other mesh clients are roaming around this area randomly.

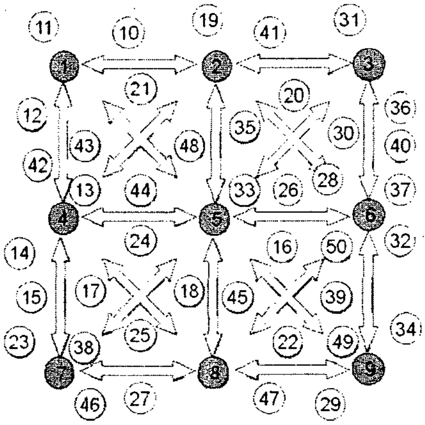


Fig. 4-7 Structure of mesh network for LAIM simulation

The simulation parameters for the first experiment are defined in Table 4-2. For each scenario, we simulated the network 30 times with different random seeds, which can result in fair accuracy. Finally, we presented the graphs with an average of 30 experiments.

Table 4-2 First Experiment Parameter for LAIM

Examined Protocols	AODV AntNet LAIM
Simulation Time	600 seconds
Network Size	800 * 800 M^2
Number of Mesh Routers	9
Number of Mesh Clients	41
Movement Pattern	Random Waypoint
Propagation Model	Two-Ray
Radio Range	298.56 M
Data Rate	2.0 Mbps
MAC Layer Protocol	802.11 b DCF
Traffic Type	CBR (UDP)
Maximum Connections	20
Data Session Start Time	Random Number in (0,80)
Data Session End Time	End of Simulation
Transmission Rate	256 Bytes/Second
Node Speed	0,5,10,20,30,40,50 M/S

For the second setting, we evaluate the protocol scalability by varying the network size, the concrete parameters used for the network topology and the size of the area are given in Table 4-3. The data traffic consists of various CBR sources sending one 256-byte packet per second. The nodes move according to random waypoint model [35], with a minimum speed of 0 m/s, a maximum speed of 15 m/s, and a pause time of 10 seconds. Each experiment is run for 400 seconds.

Table 4-3 Network Environment for LAIM

Area Size(m^2)	Clients	Routers	Total Nodes	Connections
800 × 800	41	9	50	20
1000 × 1000	64	16	80	30
1200 × 1200	75	25	100	40
1600 ×1600	151	49	200	80

4.5.2 Simulation Results

4.5.2.1 Effect of Speed on Mobility

Fig. 4-8 shows that, as the speed of nodes goes up, the PDR of all protocols degraded. The PDR of AODV decreased from 98% to 83%, which witnessed the steep change. For ACO algorithms, AntNet dropped from 99% to 89%, while LAIM has best performance in PDR, from almost 100% to 93%. From this result, it is seen that AntNet and LAIM react faster than AODV, in terms of mobility changes. This is because AODV performs relatively static routing, while both AntNet and LAIM maintain the network topology actively. LAIM provides a backup path, so it is less possible to lose the data than AntNet.

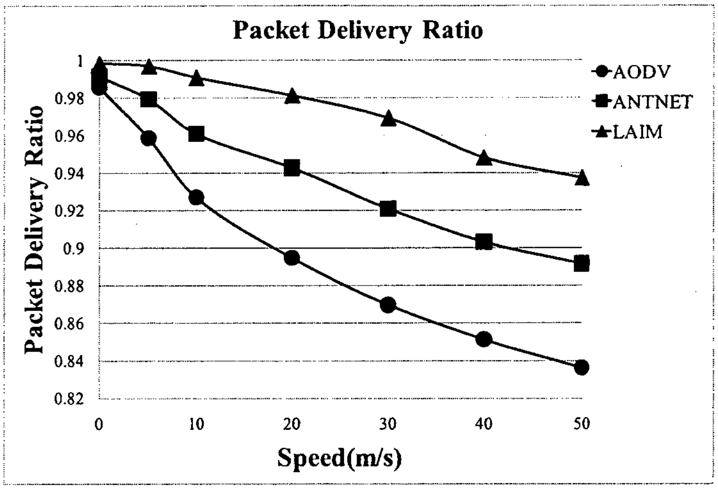


Fig. 4-8 Packet Delivery Ratio at different speeds in LAIM simulation

The total end-to-end delay in the network is compared in Fig. 4-9. As shown, the three protocols have the same trend to have more delays when speed gets higher. LAIM outperformed the others at any speed. As the speed goes up, AODV suffers from highest delay, which may be caused by its feature of replying at intermediate nodes. The routes at the intermediate nodes could be quite stale, when it comes to mobile network. Furthermore, number of hops as routing metric, can create long distance route between source and destination, which gets higher delay. AntNet uses Round Trip Time as the basic routing metric, concerning more about actual transmission time of a packet, which is more realistic. However, continuous network exploration with Ant packets will cause certain delay for the data packets in the same queue. LAIM creates specific queue for all the Ant messages, and takes into account multiple routing metric, and updates the network information periodically. It is more like the hybrid routing compared with AODV and AntNet.

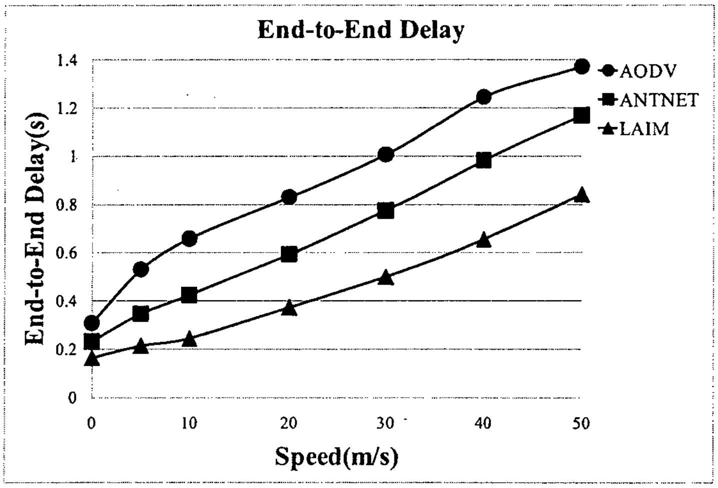


Fig. 4-9 Total End-to-End Delay at different speeds in LAIM simulation

However, ACO algorithms still have the problem of high control overhead, due to the ant features. LAIM improves a little bit in overhead ratio, compared with AntNet. As shown in Fig. 4-10, when the clients become more mobile, the overhead ratio of AODV is still within 100%, while LAIM and AntNet reach 135% and 178% respectively at the speed of 50 m/s. This is because ACO algorithms generate a slightly bigger consumption of network resources, but this is widely compensated by the much higher performance they provide, especially in terms of scalability and mobility.

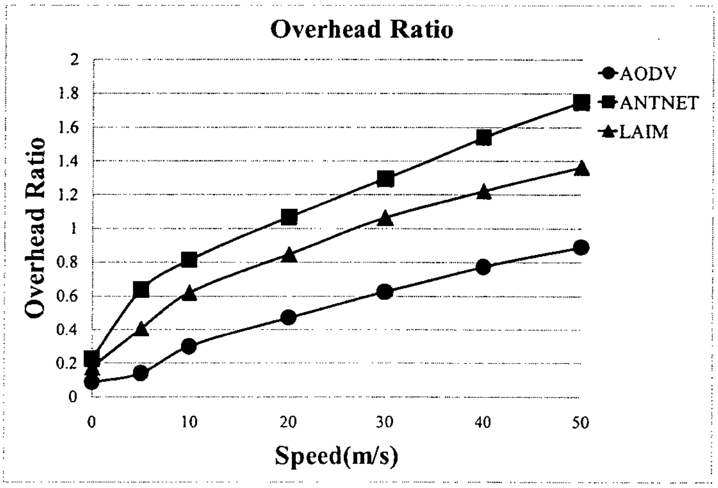


Fig. 4-10 Overhead Ratio at different speeds in LAIM simulation

4.5.2.2 *Effect of Pause Time on Mobility*

In Figs. 4-11, 4-12, and 4-13, comparisons among these three protocols are shown, in terms of various pause times. Pause time is defined as the time of staying still, when the mobile clients reach new positions. For instance, if the pause time is 0 second, then all the mobile clients keep moving until the end of the simulation. When the pause time becomes 200 seconds, all the mobile clients will stay still for 200 seconds, whenever they reach new positions. As explained, it is not difficult to see that the pause time is another factor affecting mobility of the clients, and the effect of pause time on performance is similar to that of moving speed, with slight difference.

Take Fig. 4-11 as an example, PDR of all protocols almost hit 100% together, when the pause time is 600 seconds, which is the entire simulation period. In this case, all the clients just change the initial positions, and then stay still.

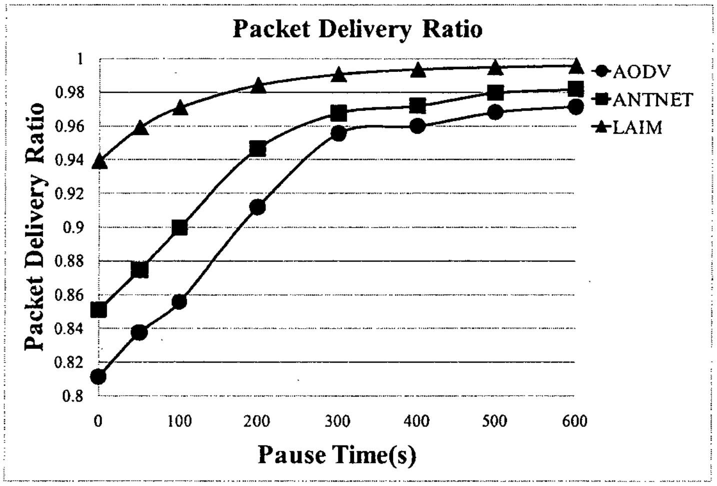


Fig. 4-11 Packet Delivery Ratio with different pause times in LAIM simulation

Similarly, for Figs. 4-12 and 4-13, as mobile clients pause for more time, which means the network becomes less mobile, end-to-end delay and overhead ratio decrease accordingly.

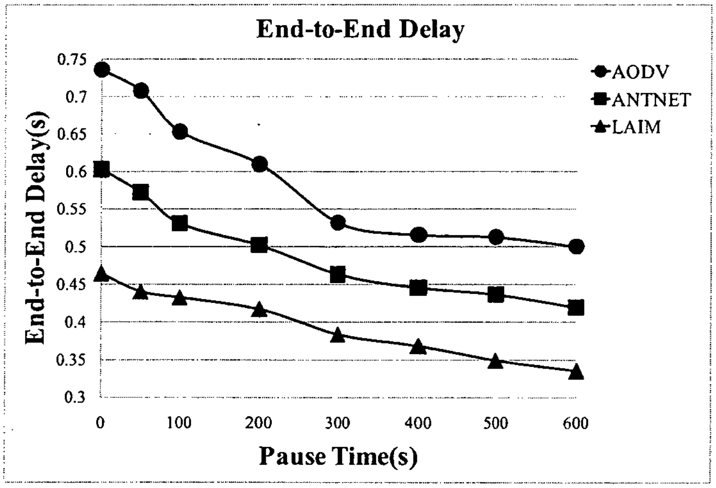


Fig. 4-12 Total End-to-End Delay with different pause times in LAIM simulation

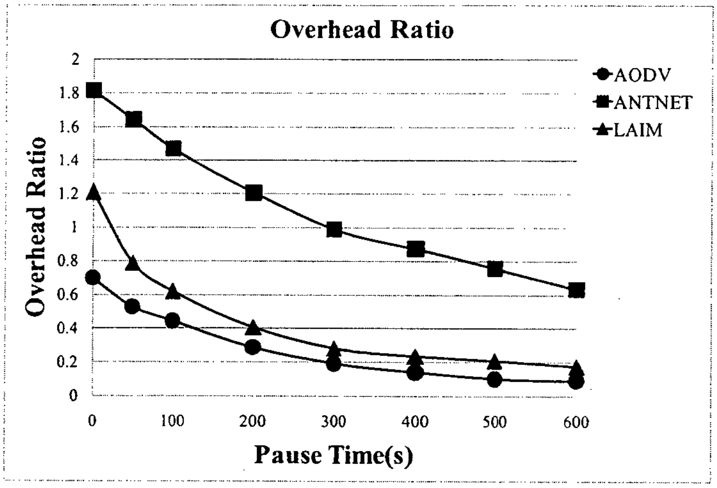


Fig. 4-13 Overhead Ratio with different pause times in LAIM simulation

4.5.2.3 Network Size

Fig. 4-14 presents the curves for packet delivery ratio as a function of network size. LAIM achieved better performance than AODV and AntNet in all network sizes.

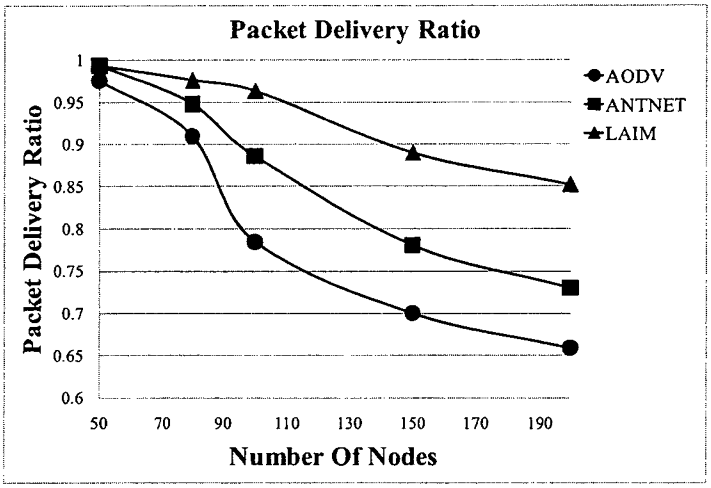


Fig. 4-14 Packet Delivery Ratio at different network sizes in LAIM simulation

With respect to total end-to-end delay in Fig. 4-15, similar results were shown.

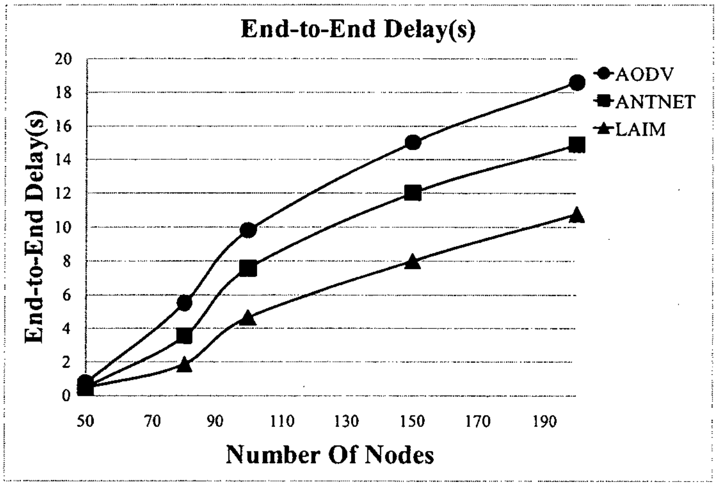


Fig. 4-15 Total End-to-End Delay at different network sizes in LAIM simulation

Due to the Ant message update, LAIM still has more overhead than AODV, while AntNet sees the highest overhead ratio in Fig. 4-16.

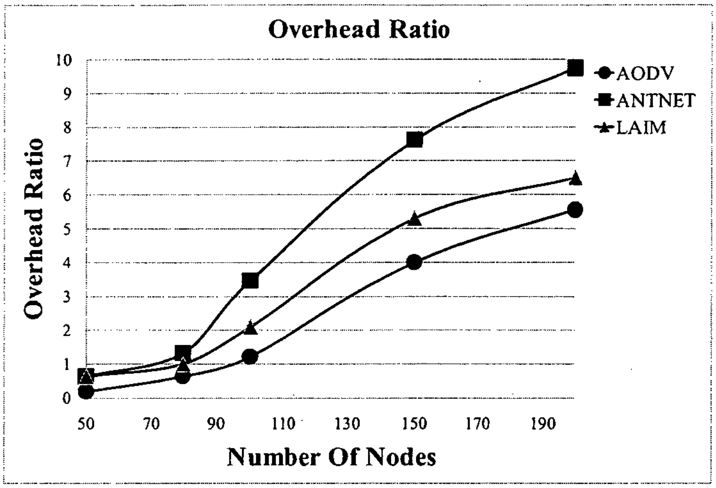


Fig. 4-16 Overhead Ratio at different network sizes in LAIM simulation

When the diameter of network increases, average number of routing hops increase correspondingly. AODV potentially may find routes with low throughput and high delay due to its features discussed above; AntNet is a special proactive routing protocol, which provides better results than AODV, however, time stamping at each intermediate node creates lots of overheads and the probabilistic entry of routing table still causes certain problem with data transmission, also forward Ants experiencing the real load may introduce extra delay and some packets loss. LAIM updates routing information only between neighbors, while the Ant collecting such information on its way to the destination, these two tasks are performed separately, and LAIM adopts proactive maintenance to keep the network refreshing.

4.6. Summary

This chapter introduces the LAIM routing protocol in detail. Like AIM, LAIM is also a hybrid routing protocol, with reactive path setup and proactive path maintenance. In LAIM, We adopt a new bandwidth estimation method named packet pair algorithm, similar approach can be seen in [17], the estimated bandwidth is more accurate in this way. For load-balancing, LAIM efficiently allocates the traffic by gathering load information from neighborhood, and by combining this information with other quality parameters, such as delay, loss ratio, and bandwidth. Thus, LAIM balances the load without generating extra overhead, as AIM does using multipath routing, which may even cause more delay and the problems of disordering and security.

Additionally, LAIM collects and exchanges network information using “Ant” message, which contains a useful parameter named ETT, it consists of packet size, loss ratio, and estimated bandwidth. Thus, network status is maximally estimated. As shown in the simulation, LAIM achieves better results than AODV and AntNet, in terms of packet delivery ratio, end-to-end delay, and overhead ratio.

CHAPTER 5

Conclusion and Future Work

5.1. Conclusion

In this thesis, we briefly investigate the literature of wireless networking, and the related problems and challenges are stated. As the wireless technology goes further, wireless mesh networking shows in the picture, as a competitive technology for next-generation networks. We introduce the structure of wireless mesh networks and their features, including their unbeatable advantages. However, there are also many challenges, such as disappointing network performance when the network increases its mobility and size, and better routing protocols are required in wireless mesh network.

Chapter two reviews current routing protocols and discusses how to design a routing algorithm under wireless environment, in order to get optimal performance. Then we present Ant Colony Optimization algorithms, which could be used in wireless routing. Moreover, we introduce the necessary concepts for designing our routing protocols, such as maximum routing metric design, load balance and bandwidth estimation, which are used in proposed routing protocols.

Chapter three and four give the details of two proposed wireless routing protocols for wireless mesh networks: Ant-In-Mesh and Load-aware-Ant-In-Mesh routing algorithms, which have many phases in common and also much difference on how they deal with network load and status. AIM uses "Hello" messages to notify the neighbors and update network status between neighbors. Since connections between the station and its neighbors can reflect load status in neighborhood, the number of connections is used to design the routing metric. AIM uses multi-path routing mechanism, which improves the overall network performance and reacts fast to data loss due to link failure.

In LAIM, more powerful bandwidth estimation method named packet pair algorithm is adopted, to estimate more accurate bandwidth. Instead of multi-path routing, LAIM filters two best paths out of the candidates, one for primary data transmission, while the other is used as backup path. Two-path routing may not deal with link failure very well, however, it is simple and problems of disordering and security are avoided, moreover, there is no extra delay or overhead existing in multi-path routing. In designing the routing metric, LAIM considers delay, bandwidth, loss ratio and transmission queue length. All these information are learned by exchanging "Ant" messages between neighbors, meanwhile, information of network status and traffic load is gathered.

Through numerous simulation experiments, we show that our proposed protocols perform well in mobile, dynamic and scaling networks, they outperform AODV and AntNet, in terms of packet delivery ratio, and end-to-end delay.

5.2. Future Work

In the following sections, we discuss two advanced topics in wireless network routing: cross-layer design and multi-radio telecommunication. These may extend our current work and provide a larger picture in this field.

5.2.1 Cross-layer Design

In traditional wired networking, the protocols are often divided into multifunctional modules at different layers, in order to establish the protocol stack. At each layer, the software takes advantage of the received information and services from the layer below, and then the service is provided to its upper layer. Data is only exchanged between adjacent layers in certain order, the communication principle is quite strict, and thus layering design and implementation are made simple and correct. In the past decades, this layering principle benefited the development of wired networks, especially the exploration of the Internet. Accordingly, Open System Interconnection (OSI) Model is becoming more and more popular.

However, things are quite different under wireless environment. Unlike the wired networks, wireless networks have relatively low link capacity and high bit error rate, due to limited frequency allocations and channel considerations. Additionally, wireless networks differ from wired ones in the following aspects:

- The wireless channel changes over both space and time, it has unstable and unpredictable characters, since normally wireless devices are moving randomly and the surrounding environment changes accordingly. Therefore, wireless communication suffers from unsuccessful transmission with errors, and the quality cannot be guaranteed.
- Within certain range, wireless transmissions may be affected by the others, collision occurs when two stations intend to occupy the same medium, thus the link capacity is limited by congestion.

Dynamic behavior is an important issue for wireless networks. But in traditional layering model, the layers are designed to operate under the worst conditions, rather than adapting to changing conditions, which leads to inefficient use of spectrum and energy [36]. Thus, conventional layering protocol stack is not sufficient for the case in wireless networks, and information sharing among different layers is necessary. In order to optimize the network performance, direct communication and variables sharing should be enabled between nonadjacent layers, meaning the structure of OSI model needs to be violated. In recent years, cross-layer design methodology is proposed to escape from the original concept of the OSI communications model with virtually strict boundaries between layers.

Basically, the key of cross layer approach is to make use of all the available information in the network, to achieve different goals, by dynamically transporting feedback regardless of the layer boundaries, enabling interaction among the layers in the protocol stack. Thus, it is essential to understand the functions of each layer:

Physical layer: raw data bits are transmitted within certain distance range, at a possible power level. Thus, the physical layer stores the available information like transmit power, bit-error rate and coding, which form the base of the network.

Link/MAC layer: the main functions of this layer are improving link reliability using Automatic Repeat reQuest (ARQ) and forward error correction (FEC), moreover, it is also required to avoid/reduce collisions, and to fragment data into frames. The available information at link/MAC layer could be corresponding FEC scheme, number of retransmitted frames, and frame length.

Network layer: routing, network addressing, network interface selection, and IP hand-off are the main functions at this layer. Accordingly, Mobile-IP hand-off initiation/completion events and the current network interface are stacked at network layer.

Transport layer: this layer aims to establish end-to-end connections over the network. As wireless networks have large delays, packet losses and high bit-error rates, cross-layer feedback is quite important at this layer.

Application layer: this layer is the interface between the network and the user, user can browse the websites, download and upload data from and to the internet, send emails, video chat with friends, etc. However, current available applications were created only for wired networks, and wireless networking is quite challenging for these applications. With cross-layer design, an application layer can provide users' quality requirements to the other layers, such as desired throughput, the delay tolerance, and acceptable packet loss rate.

Cross-layer information feedback can be carried out from upper to lower layers, and also from lower to upper layers. For instance, as soon as the application obtains the delay or loss constraints from the user, it can inform the link layer to adapt its error correction mechanisms; if the application has constraint on the throughput or any other performance, network layer can also utilize these information to design the routing protocols; among users at different consuming class, priority level can be allocated to each class at application layer, which is able to communicate with the other layers to make arrangement. On the other hand, feedback may be also transported from lower to upper layers, transmit power and bit-error rate information at physical layer can be transmitted to the link/MAC layer to enhance the error correction mechanisms; moreover, many researchers design the dynamic routing protocols based on the information from link/MAC layer.

However, current cross-layer design is still immature, any inappropriate utilization of this methodology could lead to serious problem. Normally, when cross-layer is used in wireless networking, the effect on the other layers must be taken seriously, maintain the balanced tradeoff between network performance and stability.

5.2.2 Multi-radio Routing

Although wireless networking technology is developed rapidly these years, it still suffers from significant problems in terms of scale and performance. It is concluded and proofed that overall network capacity and the end-to-end throughput of individual flows decrease rapidly as node density intensifies and the number of hops increases [37]. The most significant reason causing this is that wireless radios cannot transmit and receive data simultaneously, unfortunately, most current stations are pre-configured with only single radio. Therefore, the link capacity of relay stations is reduced seriously. Moreover, current backoff algorithms in case of congestion are not optimal enough to improve the network performance, at both MAC and transport layers. Consequently, as the network size grows, network congestion is aggravated by unbefitting algorithm, which leads to rapid degradation in throughput. Another limitation comes from the standard of IEEE 802.11, in which radios only operate over a tiny portion of available channels. The physical layer of 802.11 only allows a single frequency channel at any given time, even though there are multiple non-interfering channels available.

Aiming to solve these problems, many researchers devoted themselves to this field, and proposed several useful solutions. A number of scholars prefer to design a better MAC, which could be a very tough task, since corresponding changes need to be applied to both the software and hardware. Another way to improve wireless network capacity is to adopt rapid channel switching [38], which however, is unexpectedly slow with existing

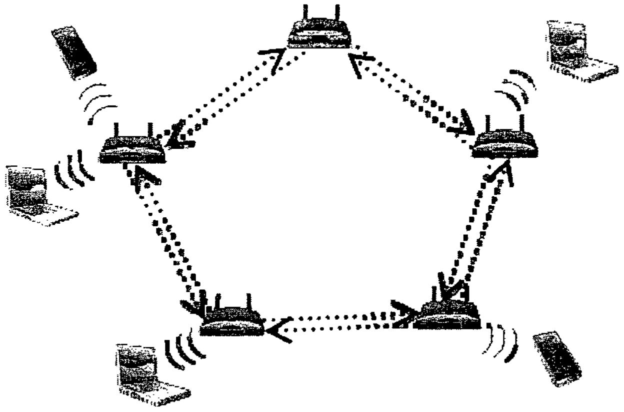
802.11 hardware.

As the cost of network interface card goes down, more and more researchers consider using multiple radios per node to improve the performance and ensure the robustness. Due to its advantageous features, multi-radio communication in the network promises a bright future for improving the capacity of these networks.

First, all the stations can transmit and receive at the same time. Otherwise, with only one radio, the capacity of relay nodes is halved and network congestion becomes an inevitable problem. Second, the radio spectrum is utilized more efficiently and completely in the network, with multiple radios, a station is able to transmit over multiple channels simultaneously. Third, there are often different types of standards in an integrated wireless network, especially in wireless mesh networks. With multiple radios enabled at each station, the entire network can operate with great compatibility.

There are many types of wireless technology existing in everyday life, such as Infrared, Bluetooth, microwave, 802.11 groups, satellite telecommunication and cellular networking technology. They are widely applied over the world, like building security system, remote control for television and other players, mobile phones, Global Positioning System (GPS), and WiMAX system. With so many technologies available, mobile devices are given more choices of access networks.

Nowadays, most handsets can connect to another headset or a laptop via Bluetooth, moreover, some mobile devices can also connect to WLAN or even satellite system via corresponding radio. Multi-radio system is an all-in-one integration of various wireless technologies, users can switch to any available radio to enjoy the service they desire. A simple multi-radio system is given in Fig. 5-1



802.11 Antenna			Bluetooth	GPS	Cellular
802.11 a	802.11 b	802.11 g			

Fig. 5-1 Multi-radio System

However, integration with multi-radio technology in wireless networks is still challenging. First, multi-function must be ensured after integrating multiple radios into the wireless device, which means this work should be stable and simple. Second, interference between multi-radio devices still exists, certain radios cannot work simultaneously, and so concurrent connections may interfere with each other, or drain power from batteries. Finally, automatic servicing of the device is quite important to the users, who require simplicity and safety. This means the wireless device needs to join and part from the network automatically, if there is any network available in the communication range.

Based on concept of multi-radio, many researchers have proposed numerous works for wireless networking. Kyasanur et al. propose a hybrid channel assignment solution for multi-radio wireless ad hoc networks [39]. In this scheme, several radios on a router are statically assigned channels, the remaining radios dynamically switch to the static channels assigned to neighboring routers in order to communicate with them. But this scheme requires that radios can only switch between channels on a per-packet basis.

Raniwala et al. [40] propose a load-aware algorithm to dynamically assign channels in a multi-radio mesh network, before channel assignment begins, the relevant traffic's information must be known. Through the simulation, network performance is improved. However, their distributed routing scheme is based on a tree rooted at a gateway, and each node has to learn 3- or 4-hop neighborhood information. Consequently, the algorithm may not work well in more dynamic networks.

Multi-radio Unification Protocol (MUP) [37] uses multiple interfaces. In MUP, when a node has K network interface cards, it only uses K channels (channels 1, 2, · · ·, K) even when there are more channels. Each node statically assigns a channel to each interface card, and when a node needs to transmit a packet, it checks the channel condition and uses the channel with the best condition at that time.

The Multi-Radio Link Quality Source Routing (MR-LQSR) protocol [17] has been developed for static community wireless networks, it assumes that the number of wireless interfaces per node is equal to the number of channels being used in the network. The protocol identifies all nodes in the wireless mesh network and assigns weights to all possible links, it is required that link information including channel assignment, bandwidth and loss rates are propagated to all nodes in the network.

In summary, the advantages of cross-layer design and multi-radio routing make them great choice to choose in designing a new routing protocol for WMN.

References

[1] M. Bahr, "Proposed routing for IEEE 802.11s WLAN mesh networks," Proceedings of the 2nd annual international workshop on Wireless internet, Boston, Massachusetts, August 2006.

[2] I.F. Akyildiz, X. Wang, W. Wang, "Wireless mesh networks: a survey," Computer Networks and ISDN Systems, vol. 47, no. 4, pp. 445-487, March 2005.

[3] J. Li, C. Blake, D. S.J. De Couto, H.I. Lee and R. Morris, "Capacity of ad hoc wireless networks," Proceedings of the 7th ACM International Conference on Mobile Computing and Networking, Rome, Italy, July 2001.

[4] M. Dorigo, T. Stützle, "Ant Colony Optimization," MIT Press, Cambridge, MA, 2004.

[5] J. Walrand, P. Varaiya, "High-performance Communication Networks," Morgan Kaufmann, 1996.

[6] H. Sandick, E. Crawley, "QoS routing working group report," Internet Draft, Internet Engineering Task Force (IEFT), 1997.

[7] C. Perkins, P. Bhagwat, "Highly dynamic destination-sequenced distance vector routing (DSDV) for mobile computers," Proceedings of ACM Special Interest Group on Data Communication, pp. 234-244, September 1994.

[8] T. Chen and M. Gerla, "Global State Routing: A New Routing Schemes for Ad-hoc Wireless Networks," IEEE International Conference on Communications, 1998.

[9] T. Clausen and P. Jacquet, "IETF RFC 3626: Optimized link state routing protocol (OLSR)," October 2003.

[10] D. Johnson, Y. Hu, D. Maltz, "IETF RFC 4728: The Dynamic Source Routing

Protocol (DSR) for Mobile Ad Hoc Networks for IPv4," February 2007.

[11] C. Perkins, E. Belding-Royer, S. Das, "IETF RFC 3561: Ad hoc on-demand distance vector (AODV) routing," July 2003.

[12] M.S. Corson, A. Ephremides, "A distributed routing algorithm for mobile wireless networks," ACM/Baltzer Wireless Networks, vol. 1, no. 1, pp. 61-81, 1995.

[13] Z. J. Haas, M.R. Pearlman, P. Samar, "The Zone Routing Protocol (ZRP) for Ad-hoc Networks," Internet Draft, July 2002.

[14] S. Radhakrishnan, N.S.V Rao, G. Racherla, C.N. Sekharan, S.G. Batsell, "DST-A routing protocol for ad hoc networks using distributed spanning trees," IEEE Wireless Communications and Networking Conference, New Orleans, 1999.

[15] D.S.J. De Couto, D. Aguayo, J. Bicket, R. Morris, "A high-throughput path metric for multi-hop wireless routing," ACM Annual International Conference on Mobile Computing and Networking, pp. 134-146, 2003.

[16] C. E. Koksal, H. Balakrishnan, "Quality-Aware Routing Metrics for Time-Varying Wireless Mesh Networks," IEEE Journal on Selected Areas in Communications, vol. 24, no. 11, pp. 1984-1994, Nov. 2006.

[17] R. Draves, J. Padhye, and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks," Proceedings of the 10th Annual International Conference on Mobile Computing and Networking, pp. 114-128, New York, NY, 2004.

[18] V. Jacobson, "Pathchar, ftp://ftp.ee.lbl.gov/pathchar/," 1997, last visited in 2009.

[19] S. Keshav, "A control-theoretic approach to flow control," Proceedings of ACM Special Interest Group on Data Communication, 1991.

[20] K. Lai, M. Baker, "Measuring Bandwidth," Proceedings of IEEE Conference on

Computer Communications, April 1999.

[21] M. Dorigo, "Optimization, learning and natural algorithms (in Italian)," Ph.D. dissertation, Dipartimento di Elettronica, Politecnico di Milano, Italy, 1992.

[22] M. Dorigo, C. Blum, "Ant colony optimization theory: A survey," Theoretical Computer Science, vol. 344, pp. 243-278, 2005.

[23] V. Maniezzo, A. Carbonaro, "Ant Colony Optimization: an Overview," Proceedings of third Metaheuristics International Conference, Brazil, 1999.

[24] D. Subramanian, P. Druschel, J. Chen, "Ants and reinforcement learning: A case study in routing in dynamic networks," Proceedings of International Joint Conference on Artificial Intelligence, Palo Alto, CA, pp. 832-838, 1997.

[25] E. Bonabeau, M. Dorigo, G. Theraulaz, "Inspiration for optimization from social insect behavior," Nature, vol. 406, pp. 39-42, July 2000.

[26] H. Matsuo, K. Mori, "Accelerated ants routing in dynamic networks," 2nd International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2001.

[27] S. Marwaha, C. K. Tham, D. Srinivasan, "Mobile agents based routing protocol for mobile ad hoc networks," Proceedings of IEEE Global Communications Conference, 2002.

[28] M. Günes, U. Sorges, I. Bouazizi, "ARA - The ant-colony based routing algorithm for MANETs," Proceedings of the International Conference on Parallel Processing Workshops on Ad Hoc Networks, 2002.

[29] G. Di Caro and M. Dorigo, "AntNet: Distributed stigmergetic control for communications networks," Journal of Artificial Intelligence Research, vol. 9, pp.

317-365, 1998.

[30] G. Di Caro, F. Ducatelle, and L.M. Gambardella, “AntHocNet: an ant-based hybrid routing algorithm for mobile ad hoc networks,” Proceedings of Parallel Problem Solving from Nature, vol. 3242, pp. 461-470, 2004.

[31] Z. Li, L. Zhishu, C. Yanhong, and X. Jianchuan, “An Ants Intelligent Routing Protocol for Wireless Mesh Network,” International Conference on Wireless Communications, Networking and Mobile Computing, 2007.

[32] M. Golshahi, M. Mosleh, and M. Kheyrandish, “Implementing an ACO Routing Algorithm for AD-HOC Networks,” International Conference on Advanced Computer Theory and Engineering, 2008.

[33] S. Lee and M. Gerla, “Split Multipath Routing with Maximally Disjoint Paths in Ad Hoc Networks,” IEEE International Conference on Communications, pp. 3201-3205, 2001.

[34] Scalable Network Technologies, Qualnet Simulator version 4.0, http://www.scalable-networks.com, last visited in 2009.

[35] D. Johnson, D. Maltz, “Mobile Computing, chapter Dynamic Source Routing in Ad Hoc Wireless Networks,” pp. 153-181. Kluwer, 1996.

[36] H. Zheng, “Optimizing wireless multimedia transmission through cross layer design,” Proceedings of IEEE International Conference on Multimeda & Expo, Baltimore, MD, July 2003.

[37] A. Adya, P. Bahl, J. Padhye, A. Wolman, L. Zhou, “A multi-radio unification protocol for IEEE 802.11 wireless networks,” International Conferences on Broadband Networks, 2004.

[38] A. Nasipuri and S. R. Das, "A multichannel CSMA MAC protocol for mobile multihop networks," IEEE Wireless Communications and Networking Conference, 1999.

[39] P. Kyasanur and N. Vaidya, "Routing and Interface Assignment in Multi-Channel Multi-Interface Wireless Networks," IEEE Wireless Communications and Networking Conference, New Orleans, LA, March 2005.

[40] A. Raniwala and T. Chiueh, "Architecture and Algorithms for an IEEE 802.11-based Multi-Channel Wireless Mesh Network," IEEE International Conference on Communications, Miami, FL, March 2005.