Protocol design and performance analysis for mobile ad hoc networks

Leng, Supeng

2005

Leng, S. P. (2005). Protocol design and performance analysis for mobile ad hoc networks. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/4592

https://doi.org/10.32657/10356/4592

Nanyang Technological University

PROTOCOL DESIGN AND PERFORMANCE ANALYSIS FOR MOBILE AD HOC NETWORKS LENG SUPENG 2005

PROTOCOL DESIGN AND PERFORMANCE ANALYSIS FOR MOBILE AD HOC NETWORKS

LENG SUPENG

SCHOOL OF ELECTRICAL & ELECTRONIC ENGINEERING

NANYANG TECHNOLOGICAL UNIVERSITY

2005

Protocol Design and Performance Analysis for Mobile Ad Hoc Networks

Leng Supeng

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirement for the degree of
Doctor of Philosophy

2005

Acknowledgements

First of all, I would like to give the deepest appreciation to my supervisor, Dr Liren Zhang, for his continuous guidance throughout the whole research project that I have been working on. Without Dr Zhang's broad vision, valuable advice and strong encouragement, this thesis would not be possible. I have not only learned enormously from Dr Zhang about the experiences of doing research, but also about the skill of communicating effectively with other researchers.

Secondly, I would like to sincerely appreciate many professors and researchers who have given me great helps in the research. Especially, the grateful thanks should give to my previous supervisor Associate Professor Krishnappa R Subramanian, Associate Professor Paramasivan Saratchandran, Professor Narasimhan Sundararajan, Professor Soon Fatt Yoon, Professor Koh Soo Ngee, Associate Professor Wende Zhong, Associate Professor Ng Chee Hock, Associate Professor Lipo Wang at Nanyang Technological University, for their encouragement and guidance in my research. Special thanks also go to all the staff in Nanyang Technological University for their effort in providing a friendly research environment.

Also, the special thanks would give to my colleagues in Communication Lab IV and Network Technological Research Centre, Dr. He Jianhua, Dr. Zheng Li, Dr. Chen Zhiyong, Dr. Ma Miao, Dr. Gu Zhenghui, Hou Fang, Huang Xiaohong, Zhang Yan, Lin Cong, Zhu Yonqing, Wu Shu and etc., who have been so patient in giving me the information I needed during my research and suggestions which were very helpful.

Finally, this thesis is dedicated to my family whose love, sacrifice and support have enabled me to mature and ultimately excel over the years – my parents, my parents-in-law and my wife.

Summary

This thesis primarily focuses on the design, modeling and performance analysis of various types of network protocols related to dynamic mobility in the network layer and the minimization of the data collisions in the Medium Access Control (MAC) layer for mobile ad hoc networks.

The first focus of this thesis is the modeling and analysis of the probabilistic characteristics of host mobility in an ad hoc network. The performances of the mobility in terms of link available time and the average number of link changes for an end-to-end connection are analyzed using mobile handover model [1] and the random walk model [2], respectively. The analysis of random walk model is especially complicated, but it is able to accurately describe the complicated host movement in an ad hoc network. On the other hand, the mobile handover model can be simply applied to characterize host mobility when hosts move at relatively higher speeds with less variation.

In this thesis, cluster structure is used to enhance network scalability. A novel k-hop clustering approach is proposed for creating more stable clusters than other existing clustering schemes. Moreover, based on the k-hop clustering structure, a novel location service protocol is presented for a large scale ad hoc network. Using an efficient inter-cluster location update mechanism, the proposed k-hop Clustering Based Location Service (KCBL) protocol is able to provide accurate mobile host location information in the neighborhood cycle. The high ability to tolerate link breakage and well reliability makes KCBL well-suited to provide efficient and robust location service and mobility management for large scale networks.

In ad hoc networks, broadcast is the only reliable delivery method when mobile hosts move arbitrarily at high speeds. Since the geometry-based broadcast protocols are more efficient than topology-based broadcast protocols for ad hoc networks with drastic host mobility, the research in this thesis aims at improving the performance of geometry-based broadcast protocols in terms of broadcast coverage. The analysis of coverage for both single-hop broadcast relay case and multihop broadcast relay case are presented. The upper bound of broadcast coverage is able to provide useful information for the design of optimal broadcast relay scheme in ad hoc networks. Such a scheme can be easily combined with the other existing geometry-based broadcast protocols for effective forwarding of packets.

The second significant focus of this thesis is to resolve the hidden terminal problem and the exposed problem [3][4] caused by large interference range in the MAC layer. With the aid of dual busy tones signaling, two approaches are proposed. The first approach is an enhancement scheme to IEEE 802.11. The second approach is to exactly signify the interference range by adjusting the transmission power of busy tones. Both approaches are able to prevent data collisions during the transmissions of data/ACK packets. The performance analysis of the proposed schemes is done using discrete Markov chain to derive the saturation throughput of CSMA MAC protocols under two-ray ground path-loss model. The performance evaluation is presented in terms of the approximate throughputs and the blocking areas. The numerical results show that the two proposed schemes outperform IEEE 802.11. Especially, the second proposed scheme is able to simultaneously resolve both of the hidden terminal problem and the exposed problem.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

Wireless networks have experienced significant evolution in recent years. One of the most rapidly developing areas is Mobile Ad Hoc NETworks (MANETs), which consists of a number of geographically distributed mobile hosts including laptop, palmtop, Internet mobile phone, sensor and etc. An ad hoc network is a self-organizing and adaptive mobile network, which is operating on mobile hosts on demand basis without specific user administration or configuration, where mobile hosts are acting as routers in a distributed manner. Compared with the other types of networks, such as cellular networks or satellite networks, the most distinctive feature of mobile ad hoc networks is the lack of any fixed infrastructure.

The concept of ad hoc network can date back to the DARPA project and multihop multiple access packet radio network (PRNET) in the 1970's [5][6][7]. However, it has received significant attention recently due to the development of wireless technology and mobile computing. In an environment lacking of fixed communication infrastructure, such as military battle field or emergency rescue area, ad hoc networks can be easily installed to provide effective communications. Moreover, wireless sensor network [8][9][10], as a special form of ad hoc networks, is widely used for digital battle field environmental or weather information gathering. In civilian areas, ad hoc networks can be deployed for commercial gatherings such as meetings, conferences, exhibitions and workshops. Bluetooth technology [11][12][13], which replaces cables by wireless connections, can be views as special kinds of ad hoc networks. New applications based on ad hoc networks have made ad hoc networks become an important part of wireless communication structure.

1.1 Motivation

When wireless devices are popularly used, ad hoc networks have the potential to simplify the deployment of the wireless network and make the communications "anywhere at anytime". Although ad hoc networks are able to provide the users great flexibility, the design of ad hoc networks faces serious challenges. The salient characteristics of ad hoc networks can be summarized as follows:

(1) Mobility: In an ad hoc network, mobile hosts are free to move arbitrarily and also mobile hosts must self-organize and reconfigure accordingly whenever they move, join or leave the network. Therefore, the network topology may change frequently and unpredictably, which makes it difficult to find suitable routes from the source to the destination. On the other hand, continuous change of mobility rates certainly increases the difficulties of the traffic control to maintain the route path across network or to switch to the alternative route when the existing link is failed. Furthermore, such mobility problems also raise the difficulties for Quality of Service (QoS) support in the network since the available resources are always varying.

(2) Multihop: When the distance between two mobile hosts in the network exceeds their radio transmission range, packets must be relayed by other inter-mobile hosts. Consequently, when network has multihop topology, packets store-and-forward routing is required across network. In conventional cellular networks or wireless networks [14][15], the wireless path is only active in the last hop between the base station and mobile user. By contrast, in ad hoc networks, all links across the network are wireless paths, which bring many challenges to the design of MAC as well as the routing

protocols. On the other hand, since mobile hosts are acting as router to store-and-forward across an ad hoc network, therefore, the security problems such as eavesdropping and spoofing have become another important issue in ad hoc networks [16][17].

(3) Contention: Comparing to the conventional cellular networks, ad hoc networks lack centralized control mechanism and global synchronization. Hence, Time Division Multiple Access (TDMA) and Frequency Division Multiple Access (FDMA) schemes are not suitable to ad hoc networks. Although the multihop topology allows spatial reuse of the wireless channels but the access to the wireless channels must be made in a distributed fashion. In this case, the nature of multihop wireless communication may result in significant contention for sharing the wireless channels in ad hoc networks. On the other hand, packet collision control over the multihop wireless channels in ad hoc networks requires numerous control overheads and unnecessary bandwidth consumption. Therefore, the MAC protocol in ad hoc networks must contend to access the channel with the capability to avoid possible packet collision between the neighboring mobile hosts.

(4) Portability: Mobile devices today are mostly operated by batteries, which are still lagging behind wireless technology. The limited power capacity of mobile hosts implies the need for power conservation [18]. On the other hand, wireless connection links in an ad hoc network have significantly lower bandwidth than those in the conventional infrastructure networks. The scarce bandwidth decreases even further due to the effects of multiple access, signal interference and channel fading.

The above mentioned technical challenges certainly affect the issues of system design and performance relate to physical layer, MAC layer, network layer and application layer including spectrum allocation, signal power control, MAC protocol design, routing protocol design, mobile host mobility management, mobile host location service, and QoS support. Since these problems in ad hoc networks are inherently difficult (NP-complete), network designers are often forced to look for suboptimal solutions. The research goal of this thesis is aimed at the modeling and performance analysis of mobile host mobility management, mobile host location service, packet broadcasting relay protocol and MAC protocol.

Unit disk graph [19] is a predominant analysis model [20][21][22] to study the topologies of ad hoc networks. The probabilistic analysis via unit disk graph is a good approximation to evaluate the performance of ad hoc networks. In this thesis, a randomly distributed ad hoc network can be modeled by a unit disk graph $G = (V, E)$, where V is the set of all the hosts in the network, and E is the set of bi-directional links. A host $h_i \in V$, or simply host i, represents the host with distinct host ID i. In a homogenous network, all the hosts have the same radio transmission range with radius r. The hosts, which are located inside of the transmission range of, h_i are called the neighboring hosts of h_i. A link (h_i, h_j) denotes a wireless link between two hosts h_i and h_i with the Euclidean distance $\left|h_i h_j\right| \le r$. Hosts find out their neighboring hosts through a Neighborhood Discovery Protocol, in which each host periodically sends a beacon to its neighboring hosts to indicate "I am alive". Two non-neighboring hosts communicate through multihop routing. In this case, the distance between hosts h_i and h_j is defined as the minimal number of hops between the two hosts.

1.2 Objectives

The key issue of an ad hoc network is the flexible host mobility, since it is able to affect the validity of routes directly. Hence, the drastic host movement may degrade network performance. The first goal of this thesis focuses on mobile host mobility analysis and the approach to reduce the effects of mobile host mobility on the network performance. A novel mobile location service protocol based on clustering structure is proposed to provide mobility management for large scale ad hoc networks. On the other hand, a broadcast relay approach is used to effectively rebroadcast packets.

The network performance in ad hoc networks are usually measured in terms of link/path throughput, transmission delay, control overheads and call blocking probability, which are highly related to host mobility. In the research works that have been done so far, simulation is used as the major tool to evaluate ad hoc network performances by taking account of host mobility [23][24][25][26][27], in which the distribution of host velocity varies as simulation progresses. Therefore, the network performance can reach steady values after a long simulation time until the simulation converges to steady state. Alternatively, the analytical derivations of mobility characteristics can directly provide stationary results to evaluate network performance, but such theoretical mobility analysis has not been used due to the complexity and flexibility of mobile host random walking movement in ad hoc networks. One task in this research is to formulate the mobility parameters in terms of link available time and the average number of link changes for host to host connections. The numerical results can be widely used in the design of ad hoc networks including traffic flow control, routing path selection, channel assigning, control overhead estimation and QoS management.

These formulas are also used as the theoretical fundamentals for the further performance evaluation in this thesis.

On the other hand, drastic host mobility is the major obstacle in the design of large scale ad hoc networks. The network routing protocol and the hierarchical network architecture based on mobile host locations are the two important technologies related to the stability and scalability of ad hoc networks. Mobile host location information has recently been applied to improve the performance of routing protocols [28][29][30][31][32], in which mobile hosts are able to obtain their own location information by using the low power low cost Global Positing System (GPS) receivers.

Due to the technology development in signal processing and VLSI (Very Large Scale Integration) design, the price of GPS chip has been dropt to only a few dollars. On the other hand, the use of Differential GPS can offer positional accuracies in a scale of a few meters. Therefore, it is reasonable to assume that each host in ad hoc network is able to be equipped with a GPS chip in the near future. However, in an indoor environment, the GPS signal may be too weak to determine location information. In this case, other available techniques [33][34] can be used to replace GPS. For example, the Time of Arrival (TOA) method [33][35] is able to obtain the distance between two mobile hosts in indoor environment. Despite the distance measurement errors and the motion of the hosts, a distributed GPS-free algorithm [33] can be applied to calculate relative coordinates and provide enough stability and location accuracy to sustain basic network functions.

Under a location based routing protocol, a mobile host must know the location information of the other mobile hosts to which the packets are going to be sent. Both the destination host

location information and the source host location information must be labeled in the packet header. For intermediate host, it needs to know the location of the neighboring hosts in order to select proper host for forwarding data packet. Typically, location service is responsible for the above tasks.

However, most location based routing protocols have assumed that the location information of mobile hosts in the network is given. In this case, the location service protocol is required to associate each individual host to find the physical location or the logical affiliation. Mobile hosts must register their current location with the server. When a host does not know the location or logical affiliation of a desired communication partner in the network, it contacts the location server for the location information. However, the location server [28][32][36][37] in an ad hoc network can only be provided in a distributed manner, since location servers are dynamically selected from the ordinary hosts equipped with the same devices.

On the other hand, hierarchical strategies [32][38][39][40][41] are used to construct large diameter ad hoc networks. The most popular way of building hierarchy is to group hosts geographically close to each other into clusters [39][40][41]. Cluster architecture makes a large network appear much smaller, and make a highly dynamic topology appear much less changing. Each cluster has a cluster-head to manage the membership on behalf of the cluster. An efficient clustering scheme tends to keep cluster-heads as stable as possible, and preserve clustering structure when a few hosts are moving. Otherwise, high processing and communication overheads should be paid to reconstruct the clusters. Since the cluster size of k-hop cluster ($k > 1$) is more than that of 1-hop cluster, k - hop cluster structure has more scalability. A novel k-hop Compound Metric Based Clustering (KCMBC) approach is proposed to create stable cluster structure in a dynamic environment.

Since the cluster-head acts as a local coordinator within the cluster, the instinctive and best role of location server is the cluster-head itself. Our proposed location service protocol is based on k -hop clustering structure. The cluster-heads works as distributed location servers, which maintains the inter-cluster connectivity and the membership of each cluster. The location enquiry from a cluster member can be immediately responded by its cluster-head. As routing is still operated in a flat network structure, the communication bottlenecks at cluster-heads are avoided and the reliability of location service is increased. Moreover, the inter-cluster location update mechanism well balances the trade off between the communication overheads and the accuracy of location information. With good stability and scalability, our proposed k -hop Clustering Based Location Service (KCBL) protocol is able to provide location service for most routing protocols.

Because of the ever-changing topology, ad hoc networks use broadcasting as one of the fundamental protocols. Especially when packets are transmitted to multiple hosts, broadcast is more efficient than unicast or multicast [42][43]. It is more important that the broadcast is the only reliable delivery method when mobile hosts move arbitrarily at a high speed. There are many routing protocols [44][45][46][47][48] that rely on broadcast to discover routing information and implement routing recovery. Current research on optimal broadcasting mechanisms in ad hoc networks has been focusing on minimization of the number of broadcasts and the efficiency of delivering packets to the other hosts in the network. In this thesis, the research aims at improving the performance of geometry-based broadcast protocols [49][50][51][52], which choose the forwarding hosts according to the geometry location information of the neighboring hosts. The upper bound of broadcast coverage area

can be used to design a novel broadcast relay scheme, which can help geometry-based broadcast protocols to achieve higher delivery rate and lower number of broadcasts.

Medium access control is another active research area for ad hoc networks. Although IEEE 802.11 MAC [14][53] is widely used in ad hoc networks, the RTS/CTS handshake and virtual carrier sensing function contained in IEEE 802.11 MAC are not effective to be used for the estimation of interference in ad hoc network environment. This is because that hidden hosts cannot hear RTS/CTS dialogues when the interference range becomes large (the detail can be found in Section 6.2). In the second part of the thesis, two novel MAC schemes are proposed to help IEEE 802.11 to combat the hidden terminal problem and the exposed problem [3][4] caused by large interference range. With the aid of dual busy tones, one scheme is a simple enhancement to IEEE 802.11, and the other is to exactly signify the interference range by adjusting the transmission power of busy tones. Both the schemes are able to prevent all collisions during the transmissions of data/ACK packets. The proposed Variable Power Dual Busy Tone (VPDBT) MAC balances the trade off between spatial reuse and collision avoidance. Furthermore, using discrete Markov chain, we present an analytical model to derive the saturation throughput of the CSMA protocols under two-ray ground path-loss model.

1.3 Major Contributions of the Thesis

- First of all, probabilistic analytical models are proposed to formulate the stochastic characteristics of a homogenous ad hoc network with respect to: (1) the average number of neighboring hosts of each host, (2) the average number of hops for an arbitrary communication pair, (3) the link available time and (4) the average number

of path/link changes for each connection. Especially, the latter two mobility related parameters are derived using mobile handover model [1] and the random walk model [2]. The formulas are derived using the relative velocity between neighboring hosts, which is assumed to follow a general distribution during each moving epoch. Moreover, the study focuses on the performance of mobility in a specific case where the host random moving including speed and direction is assumed to have a uniform distribution, which has commonly been used for investigations on ad hoc networks. Although the random walk analytical model is more complicated, it can better describe the complicated host movement in an ad hoc network. The analysis based on random walk model can better evaluate the performance of network than that based on the mobile handover model. Furthermore, the mobile handover model can be simply applied to characterize host mobility when hosts move at relatively fast speeds, or hosts change their speeds slowly.

- A novel k -hop Compound Metric Based Clustering (KCMBC) approach is proposed, which not only inherits the scalability of k -hop clustering for large diameter networks, but also uses some innovations to improve the performance of clustering: (1) the average link expiration time of each host is used to characterize the degree of mobility of a host with respect to its neighboring hosts; (2) only the cluster-head candidates, which satisfy the condition of host mobility, can be elected as cluster-heads; (3) the compound metric combining host connectivity, the average link expiration time and host ID is used to select cluster-heads; (4) the distance related converge-cast takes the place of gateway initiated converge-cast to gather the information of cluster members. (5) an efficient cluster maintenance mechanism is employed to cope with local topology changes. Simulation experiments show that

clusters created by the KCMBC approach have modest and uniform cluster size. With respect to other clustering schemes, KCMBC can increase the cluster-head duration significantly.

- A new k-hop Clustering Based Location Service (KCBL) protocol is proposed. Under KCBL, cluster-heads are dynamically distributed and assigned to perform the location server function. Cluster ID or Cluster-head's coordinates are used to represent the location of all members in the cluster. A source host can obtain the location information of the destination by simply sending one location enquiry to its cluster-head. Considering the distance effect [54] and host mobility, an efficient inter-cluster location update mechanism largely reduces the communication overheads. KCBL provides more accurate location information in the destination's neighborhood and less accurate information for the hosts far away. The frequency of inter-cluster location updates can be determined by the mobility pattern of this cluster or the group characteristics. All of these above features ensure that KCBL has good scalability. In addition, due to the benefits of the cluster structure and distributed cluster-heads, KCBL is able to increase the performance of routing protocols, in terms of overhead reduction and route recovery. With good scalability and the ability of self-organization, KCBL is able to adapt to accommodate most applications of MANET.

- The performances of the KCBL protocol are evaluated in terms of control overheads, cost for location management, and the accuracy of location enquiry response. Simulation results indicate that the total cost of location management in KCBL is much less than the cost of a link state protocol. Both the overhead in the initial stage

and the total cost of location management reduce with the increase of k. A large value of k can not only suppress the increase rate of the total cost while the number of hosts in the network increases, but also increase the hit probability of location service, and reduce the passive effect of host mobility on control overhead and the hit probability. Moreover, a suitable value of scaling factor σ can balance the trade off between the cost and the accuracy of location service.

- The analysis of the coverage area for both single-hop broadcast relay case and multihop broadcast relay case are presented. The upper bound of broadcast coverage is obtained, which is used for the design of a novel Forwarding Host Selection (FHS) broadcast relay scheme. Differing from common geometry-based protocols, the FHS scheme is based on both the relative distance and the forward angle information of neighboring hosts. The hosts within the symmetrical area have higher priority to rebroadcast the packet than other hosts. With little extra overhead and computational load, FHS can help many geometry-based broadcast protocols achieve high delivery rate and low number of rebroadcasts. Simulation shows that the broadcast performance improved by the FHS scheme increases when the network host density increases. The geometry-based protocols combined with FHS work reliably for highly dynamic ad hoc networks.

- In the MAC layer, the research focuses on the performance of IEEE 802.11 in the open space environment with large interference. Based on dual busy tones, two novel MAC schemes are proposed to reduce most interference under two-ray ground path-loss model. Both the schemes are able to eliminate all conflicts during the transmissions of data/ACK packets. Among them, Fixed Power Dual Busy Tone

(FPDBT) MAC is a simple enhancement to IEEE 802.11 DCF, and Variable Power Dual Busy Tone (VPDBT) MAC use busy tone sensing to replace the virtual carrier sensing and physical channel sensing implemented in IEEE 802.11. Using VPDBT MAC, the dual busy tones can exactly cover the interference range of a transmitter-receiver pair. Hence, VPDBT has optimal spatial reuse, and it can simultaneously solve the hidden terminal problem and the exposed problem.

- One analysis model is presented to derive the saturation throughput of CSMA MAC protocols under two-ray ground path-loss model in ad hoc networks. Moreover, we formulate the approximate throughputs and the blocking areas for IEEE 802.11, FPDBT and VPDBT. Numerical results show that VPDBT has the best performance, followed by FPDBT and IEEE 802.11. Both the VPDBT scheme and the FPDBT scheme can enhance the performance of IEEE 802.11 DCF

1.4 Organization of the Thesis

This thesis is organized as follows. Chapter 1 is the introduction that presents the context and motivation of research topic including (1) the major characteristics of ad hoc networks, (2) highlight the major contributions of this thesis, including mobility analysis, the design and performance analysis of a novel clustering scheme, location service protocol, new broadcast relay scheme, and MAC schemes for IEEE 802.11 enhancement.

Chapter 2 concentrates on the analysis of host mobility in an ad hoc network. Combining the relative velocity between two neighboring hosts, the mobile handover

model and the random walk model are deployed to derive the characteristics of random host mobility, in terms of the link available time and the average number of path/link changes for an end-to-end connection. Moreover, two important network parameters, namely the average number of neighboring hosts of each host and the average number of hops for an arbitrary communication pair, are formulated in this chapter.

Chapter 3 presents a novel KCMBC approach, which overcomes the drawback of Max-Min heuristic [55] while keeping the same fast convergent speed and good scalability. Performance evaluation indicates KCMBC is able to create more stable and more uniform cluster structure.

Chapter 4 proposes a new KCBL protocol, which is able to provide fast response for location enquiry. The details of the location information update are described, and the benefits that KCBL brings to routing protocols are also discussed. The performances of the KCBL protocol are evaluated in terms of control overheads, cost, and the accuracy of location service. The KCBL protocol is able to adapt the frequency of location updates in an ad hoc network with variable host mobility, and balances well the trade off between control overheads and the accuracy of location service.

Chapter 5 presents the conditions to achieve the upper bound of broadcast coverage in an ad hoc network, which is useful for the design of suboptimal broadcast relay scheme. Secondly, a novel FHS broadcast relay scheme is proposed, which is able to forward packets effectively, especially when network host density is heavy. Simulation experiments on delivery ratio and the number of forwarding hosts show FHS is able to enhance the performance of geometry-based protocols

Chapter 6 focuses on the MAC design for improving the performance of IEEE 802.11 with large interference range. Two novel MAC schemes named FPDBT and VPDBT are proposed. In order to compare the performances of IEEE 802.11 and the two enhanced versions, this chapter presents the first analysis model to derive the saturation throughput of CSMA MAC protocols under two-ray ground path-loss model in ad hoc networks. Through the performance evaluations in terms of blocking area, saturation throughput and aggregate throughput, it can be found that our two proposed MAC schemes outperform the original IEEE 802.11, and VPDBT is able to solve the hidden terminal problem and the exposed problem simultaneously.

Conclusion of the thesis and suggestions for future research are presented in Chapter 7. It summarizes the proposed protocols and schemes of this thesis as well as highlights the performance results and suggestions.

Chapter 2 Mobility Analysis of Mobile Hosts with Random Walking in Ad Hoc Networks

2.1 Introduction

Host mobility is one of the most important functions of routing protocols in support of ad hoc networks. Since an ad hoc network works in dynamic environment, the mobility parameters [20][56][57][58][59][60] are usually considered in traffic flow control, routing path selection, mobile channel assigning, control overhead estimation and QoS management. For example, the link available time can be used to schedule wireless medium access, to assign the exchange rate of "Hello" packets among neighboring hosts and to arrange location update in location service protocols. Because of flexible host movements, the analysis of mobility characteristics for MANET is a very challenging work. In many literatures, these parameters are assumed to be given, although this assumption is impossible due to lacking of corresponding analytical models. Hence, this chapter focuses on formulating mobility parameters, in terms of link available time and the average number of link changes for each connection.

On the other hand, host mobility also plays a critical role in the design of ad hoc networks including estimation of offered traffic load, dimensioning of signaling overhead, estimation of channel holding time and user location updating. However, each host in an ad hoc network moves with independent velocity and direction. There is no fix point of reference in this dynamic environment, which is unlike in cellular networks where the mobility of hosts is relative to a fixed base station. Hence, the mobility model

used in cellular networks cannot be applied to ad hoc networks directly. By considering the relative velocity between two neighboring hosts, the mobile handover model [1] and the random walk model [2] are deployed in this chapter to derive the random movements of hosts. Although the random walk model is more complicated, it matches the ad hoc network environment and is able describe the movement of mobile host more accurately. On the other hand, with slight modification, the probabilistic model of the random walk mobile profile can be also applied in the mobility analysis for cellular networks or clustering structures in ad hoc networks.

A number of mobility models [2][61][62] are applied in simulation and analytical studies of ad hoc networks and cellular networks. Mobility models in cellular networks are used to determine the distribution of the occupancy time for a mobile in a fixed cellular. Hong and Rappaport [1] have investigated the cellular mobility. In their works, the cell residence time distributions in the mobile handover model is derived under the condition of mobile hosts moving in random direction at a constant speed with uniform distribution. This derivation has been commonly used in the study of handover problems. Zonoozi [63] investigated the performance for the random trajectory of mobile node using simulation. His results show that gamma distribution may provide a good approximation for the cell residence time. The cell residence time of cellular system, labeled as the mobile sojourn time, is also investigated by other researchers [64][65].

However, there is no report of research on link available time and other relevant mobility parameters for ad hoc networks. Unlike cellular system in which mobility is measured relative to fixed base stations, the mobility analysis in ad hoc networks

becomes more complicated because both the transmitter and the receiver of a link are mobile. In [66], the two-body-mobility problem in wireless ad hoc networks is reduced to one-body-mobility problem by introducing relative position and motion. In this literature, the probability of link availability is analyzed with the assumption of Rayleigh distributed random mobility vector, and the mobility model provides the basis for dynamically grouping hosts into clusters.

This chapter studies the mobility performance based on the relative velocity between two neighboring hosts. Firstly, this chapter focuses on the performance analysis of ad hoc mobility that hosts are randomly moving at speeds with general distribution. Secondly, the study focuses on the performance of mobility in a specific case where the host random moving including speed and direction is assumed to be uniform distribution, which has commonly been used for the investigations of ad hoc networks.

In this chapter, both the mobile handover model [1] and the random walk model [2][65][66][67] are applied to evaluate the mobile characteristics in ad hoc networks. The random walk model is memoryless mobility pattern and a widely used mobility model for wireless networks. This model is equivalent to the random waypoint model without pause times, which is a popular mobility model used in simulation experiments [27][68][69] for ad hoc networks. Therefore, the analysis results can be migrated to some cases of the random waypoint model. The derivation methods in this chapter can be used to formulate the characteristics of other mobility models for ad hoc networks.

2.2 Relevant Technological Terms

In order to formulate the mobility parameters for ad hoc networks, two relevant static parameters of ad hoc networks, called the average number of neighboring hosts of a host (the average host degree d in the network) and the average number of hops H_a for arbitrary communication pair, are considered.

2.2.1 The Average Host Degree

It is supposed that N mobile hosts are uniformly distributed in an ad hoc network. Let p_0 denote the probability that a particular host is the neighboring host of another host [1]. Since the connectivity between two arbitrary hosts is identically, independently, distributed (i.i.d.), the probability that n neighboring hosts are located inside of the transmission range of a particular host is given by

$$p_n = \binom{N-1}{n} p_0^n (1-p_0)^{N-1-n}$$

The average degree of each host can be obtained as

$$d = \sum_{n=0}^{N-1} n p_n = \sum_{n=1}^{N-1} \frac{(N-1)!}{(n-1)!(N-1-n)!} p_0^n (1-p_0)^{N-1-n}$$

Let $m = n-1$, $M = N-2$, then

$$d = p_0 \sum_{m=0}^{M} \frac{(N-1)M!}{m!(M-m)!} p_0^m (1-p_0)^{M-m}$$

Using Binomial Theorem, then the average degree in the network is given by

---

[1] p_0 can be given by $p_0 = \pi r^2 / S$, where r is the transmission range of each host, and S is the area of the network.

$$d = p_0(N-1)(p_0+1-p_0)^M = p_0(N-1) \quad (1)$$

2.2.2 The Average Number of Hops for a Communication Pair

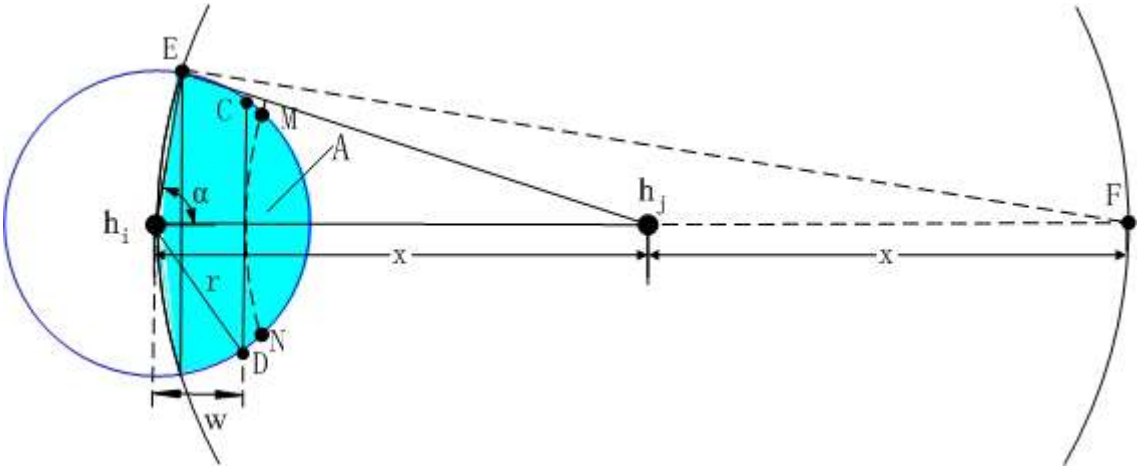


Figure 2-1. The progress distance of one hop

This subsection derives the average number of hops for arbitrary communication pair in an ad hoc network. As shown in Figure 2-1, the radio transmission range of each host is a circular area of radius r. Host h_i has packets to be delivered to host h_j which is located outside of h_i's transmission range. The distance between h_i and h_j is represented by x. In order to send packets to h_j, h_i usually selects a host located in the shadow for forwarding packets. When $x = |h_i h_j| \gg r$, the area S_A of the shadow area A can be approximated as:

$$S_A \approx \pi r^2 \frac{2}{2\pi} = r^2\alpha ,$$

where $\alpha = \angle E h_i h_j$. Since the line segment $\overline{h_i F}$ is the diameter of the circle which is centered at h_j with the radius x, $\angle F E h_i$ is a right angle and $\alpha = arccos\frac{r}{2x}$. Therefore,

$$S_A \approx r^2 \cdot arccos\frac{r}{2x}$$

In Figure 2-1, the arc $\overset{\frown}{MN}$ is centered at h_j and inside of the shadow area A. Let w denote the shortest distance from h_i to the arc $\overset{\frown}{MN}$. Since all hosts located on arc $\overset{\frown}{MN}$ have the same progress distance w [70] with respect to h_j, and $\left|\overset{\frown}{MN}\right| \approx |CD| = 2\sqrt{r^2 - w^2}$, then the Probability Density Function (PDF) of w is given by

$$f_W(w) = \begin{cases} \dfrac{\left|\overset{\frown}{MN}\right|}{S_A} \approx \dfrac{2\sqrt{r^2 - w^2}}{r^2 \cdot \arccos\dfrac{r}{2x}}, 0 \le w \le r \\ 0, \qquad \text{elsewhere} \end{cases}, \tag{2}$$

and the Cumulated Density Function (CDF) of w is given by

$$F_W(w) = \int_{-\infty}^{w} f_W(t)dt \approx \begin{cases} \dfrac{w\sqrt{r^2 - w^2} + r^2 \arcsin\dfrac{w}{r}}{r^2 \cdot \arccos\dfrac{r}{2x}}, 0 \le w < r \\ 1, \qquad w \ge r \end{cases}$$

The mean of w is $E(w) = \int_{-\infty}^{\infty} w f_W(w)dw = \dfrac{2r}{3\arccos\dfrac{r}{2x}}$

Assuming that hosts are randomly located in an ad hoc network with uniform distribution, then the average number of neighboring hosts located in the shadow area A is given by

$$d_A = \frac{S_A}{\pi r^2} d \approx \frac{d}{\pi} \cdot \arccos\frac{r}{2x},$$

where the average degree of a host can be derived by equation (1). The nearest integer no larger than d_A is denoted as $d_{AI} = \lfloor d_A \rfloor$.

For a mobile host h_i in an ad hoc network, when it selects the next hop to deliver packet, it usually selects the neighboring host that has nearest distance to the destination h_j [71][72][73], i.e., the neighboring hosts that are located in the shadow area A with the

maximum progress distance to h_i. Let w_n (n = $1,2,\cdots d_{AI}$) denote the progress distance of each neighboring host of h_i, with the same distribution expressed by equation (2). Since the location of hosts is independent, the maximum progress distance w_A can be given by

$w_A = Max(w_1, w_2, \cdots, w_{d_{AI}})$,

The CDF of w_A can be obtained as

$$F_{W_A}(w_A) = \prod_{n=1}^{d_{AI}} F_{W_n}(w_n) = F_W^{d_{AI}}(w)$$

The PDF of w_A is given by

$$f_{W_A}(w) = \frac{dF_{W_A}(w)}{dw} = d_A F_W^{d_A - 1}(w) f_W(w)$$

$$= \begin{cases} \dfrac{2d_A\sqrt{r^2-w^2}\left(w\sqrt{r^2-w^2} + r^2 arcsin\dfrac{w}{r}\right)^{d_A-1}}{\left(r^2 \cdot arccos\dfrac{r}{2x}\right)^{d_A}}, 0 \le w \le r \\ 0, \qquad \text{elsewhere} \end{cases}$$

The expected mean value of w_A can be obtained as

$$E(w_A) = \int_{-\infty}^{\infty} w f_{W_A}(w) dw = \frac{2d_A \int_0^r w\sqrt{r^2-w^2}\left(w\sqrt{r^2-w^2} + r^2 arcsin\dfrac{w}{r}\right)^{d_A-1} dw}{\left(r^2 \cdot arccos\dfrac{r}{2x}\right)^{d_A}} \quad (3)$$

Let l_a denote the distance between two arbitrary points located within the square area $[0,l]^2$. According to [70], the average number of hops for an arbitrary communication pair can be approximated as

$$H_a \approx \frac{E(l_a)}{E(w_A)} \approx \frac{0.261l\left(r^2 \cdot arccos\dfrac{r}{1.043l}\right)^{d_A}}{d_A \int_0^r w\sqrt{r^2-w^2}\left(w\sqrt{r^2-w^2} + r^2 arcsin\dfrac{w}{r}\right)^{d_A-1} dw} \quad (4)$$

where $d_A \approx \left\lfloor \frac{d}{\pi} arccos \frac{r}{1.043l} \right\rfloor$.

The expected mean of l_a can be expressed as the quadruple integral

$$E(l_a) = \frac{1}{l^4} \int_0^l \int_0^l \int_0^l \int_0^l \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}\, dx_1 dx_2 dy_1 dy_2 \tag{5}$$

According to [74][75], $E(l_a)$ is given by

$$E(l_a) = \left[\frac{2+\sqrt{2}}{15} + \frac{\ln(1+\sqrt{2})}{3} \right] l \approx 0.5214l$$

In equation (4), when l is a fixed value, H_a is determined by the transmission range r and the average degree d of each host. Figure 2-2 shows the curve of H_a in a square area with $l = 10$ units. It is clear that H_a largely decreases with the increase of r, and the increase of d results in the decrease of H_a.

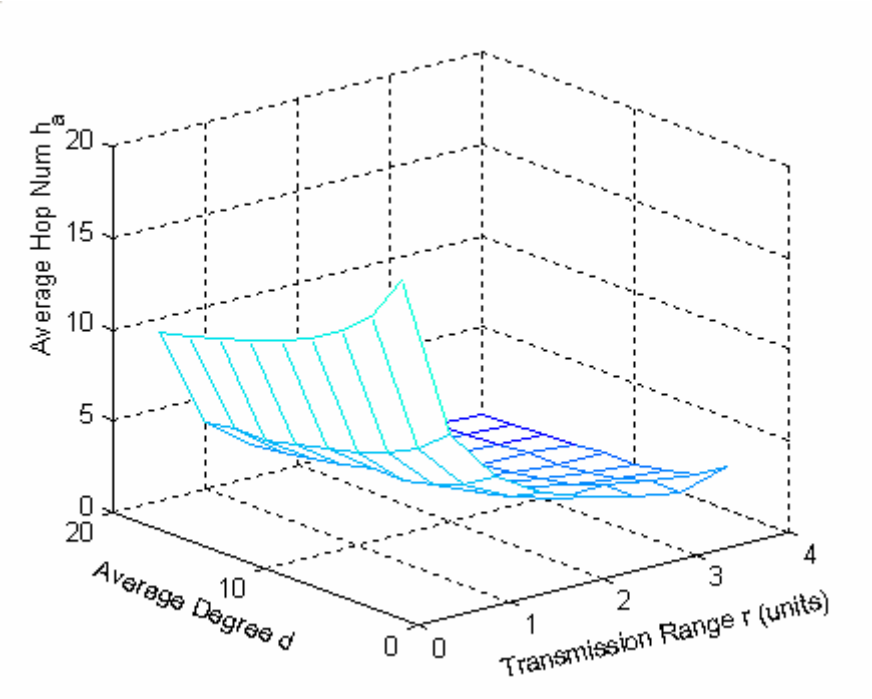


Figure 2-2. Average number of hops H_a, $l = 10$ units

2.3 Modeling and Performance Analysis for Network Mobility

The ad hoc mobility modeling proposed in this section is a continuous-time stochastic process, which characterizes the movement of mobile hosts in a two-dimensional space. The two-body-mobility problem in wireless ad hoc networks is reduced to single-host mobility problem by introducing the relative velocity between two neighboring hosts.

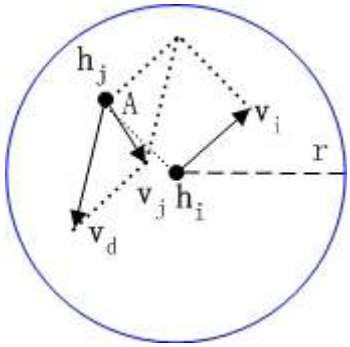


Figure 2-3. Relative velocity vector between two hosts

As shown in Figure 2-3, the instantaneous moving of two neighboring hosts h_i and h_j are represented by velocity vector $\mathbf{v}_i$ and $\mathbf{v}_j$, respectively. In order to simplify the modeling, h_i is located at the center point of the "cell" which has a radio transmission range of radius r, as a reference point, while h_j is moving in the range of the "cell". Therefore, the instantaneous velocity difference between h_i and h_j can be represented using a relative velocity vector $\mathbf{v}_d = \mathbf{v}_j - \mathbf{v}_i$. If the moving rates of these two vectors $\mathbf{v}_i$ and $\mathbf{v}_j$ are all in a range of $(0, V_m)$ with a general distribution, where V_m is the maximum moving rate, then the moving rate of velocity $\mathbf{v}_d$ is obviously in the range of $(0, 2V_m)$. In this case, the corresponding mobility of these two hosts can be modeled as the problem whether h_j

remains in the "cell" of h_i. On the other hand, the distribution of link available time between h_i and h_j is able to be calculated using the PDF of relative velocity vector $\mathbf{v}_d$.

2.3.1 Host Sojourn Times

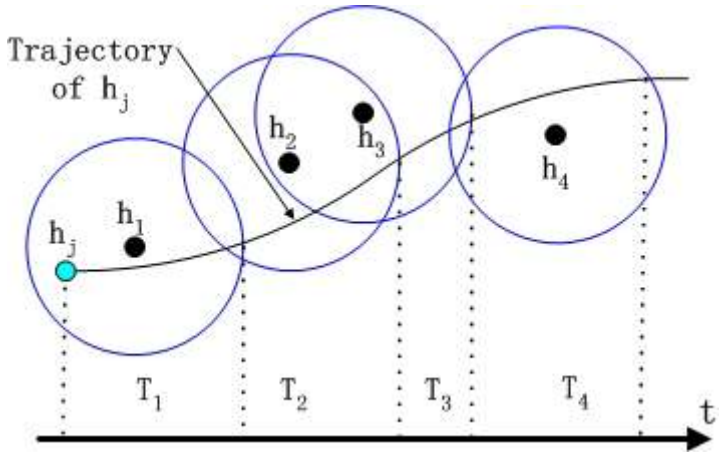


Figure 2-4. The sojourn time for host h_j traveling across conjoint hosts

Figure 2-4 illustrates that host h_j travels through conjoint hosts h_i $(i = 1, 2, 3, \ldots)$ while h_j keeps an end-to-end connection to communicate with its destination. At any time, h_j is located in the transmission range of at least one host, and h_j is always able to switch the communication link to the conjoint hosts to maintain an active link when the existing link is broken. For example, when h_j moves out from the cell of h_1, the handover occurs that a new wireless link between h_j and h_2 is established to replace the previous link between h_j and h_1. Accordingly, an alternative path will be discovered to keep the on-going communication between h_j and its destination. As shown in Figure 2-4, T_i is the link available time of the wireless link between h_j and individual h_i $(i = 1, 2, 3, \ldots)$. In this case,

the link available time is defined as the sojourn time T_X, which is similar to the remaining sojourn time [64] in the cell for call initialization in cellular system

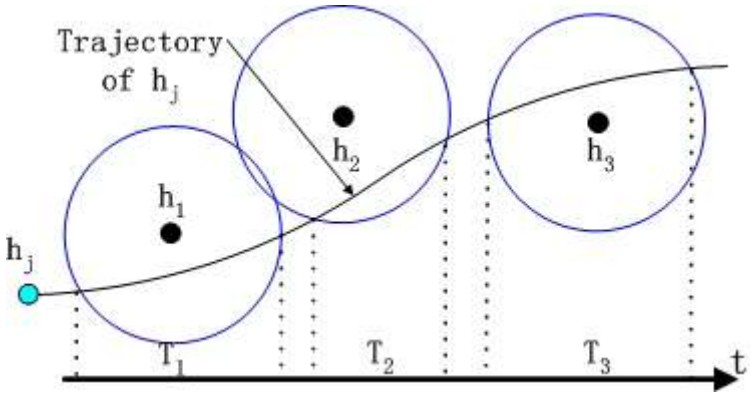


Figure 2-5. The sojourn time for host h_j traveling across disjointed hosts

On the other hand, as shown in Figure 2-5, if the host density in an ad hoc network is low, then mobile host h_j cannot connect to any other host when h_j leaves from the cell of its previous neighboring host h_i ($i = 1, 2, 3, \ldots$). The end-to-end connection between h_j and the corresponding destination has to be terminated before the communication task is completed. In this case, the link available time of the wireless link between h_j and individual h_i ($i = 1, 2, 3, \ldots$) is named as the sojourn time T_S, which is equivalent to the sojourn time in an intermediate cell of cellular network [64].

Combining the host sojourn time T_X and T_S, the average link available time is given by

$$T_{HAV} = \xi E(T_X) + (1-\xi) E(T_S),$$

where $E(T_X)$ and $E(T_S)$ are the expected mean values of the sojourn time T_X and the sojourn time T_S, respectively, and ξ is the probability that the host h_j is located inside of

the cell of another neighboring host at the moment when h_j leaves from the transmission range of h_i.

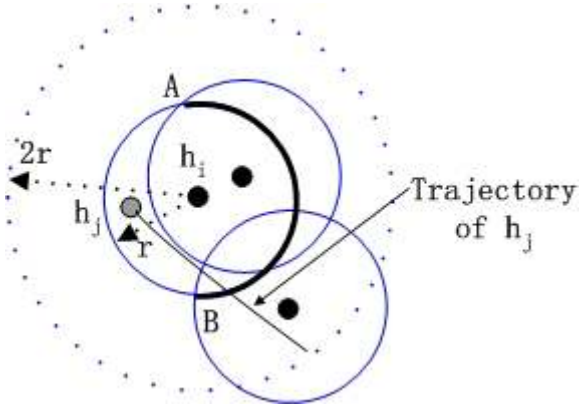


Figure 2-6. The average length of arc $\widehat{AB}$

As shown in Figure 2-6, the hosts that are located in the circular area of the center h_i with the radius $2r$, are defined as the conjoint hosts of h_i. Then, ξ can be derived by calculating the average length of arc $\widehat{AB}$[2], which is the arc on the boundary of h_i's transmission range and inside of the transmission ranges of h_i's conjoint hosts. The average number of conjoint hosts around h_i is given by

$$d_{hj} = \frac{\pi(2r)^2}{\pi r^2} d = 4d \tag{6}$$

If $d_{hj} = 1$, the average length of $\widehat{AB}$ can be obtained as

$$E\left(\left|\widehat{AB}\right|\right) = \int_0^{2r} \left|\widehat{AB}\right| \frac{2\pi x}{4\pi r^2} dx = \frac{1}{2r^2} \int_0^{2r} 2rarccos\frac{x}{2r} \cdot x dx = \frac{\pi r}{2}$$

Therefore, $\xi = \dfrac{E\left(\left|\widehat{AB}\right|\right)}{2\pi r} = \dfrac{1}{4}$

[2] The arc may consist of a few portions if more than one conjoint hosts are located around h_i

However, when $d_{hj} > 1$, the conditions of overlapping arcs among conjoint hosts become very complicated and the formula derivation of $E\left(\left|\widehat{AB}\right|\right)$ is difficult as well. In this case, simulation is used to get the numerical results of ξ, which is shown in Figure 2-7.

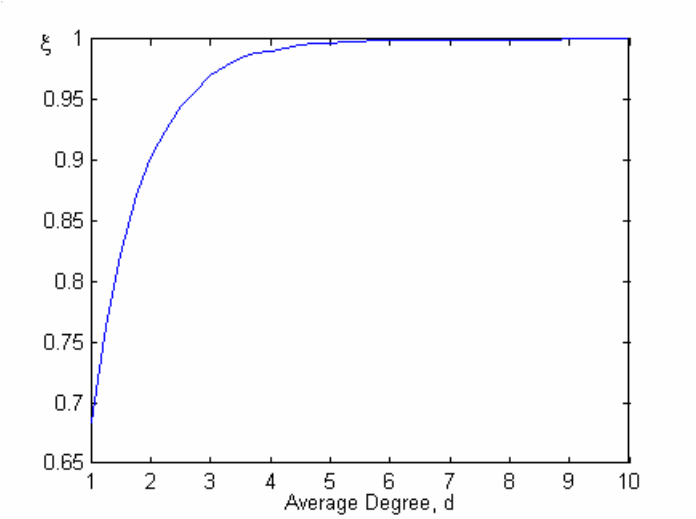


Figure 2-7. ξ versus host degree

Figure 2-7 illustrates that when $d \geq 3$, the corresponding value of $\xi > 0.95$. It is obvious that the value of T_{HAV} is significantly determined by $E(T_X)$, especially for high host density cases. On the other hand, it is clear that the moving scenario shown in Figure 2-5 is impossible to provide end-to-end connections in an ad hoc network. Therefore, the following analysis presented in this chapter only focuses on the case that a mobile host is always able to switch the communication link to the conjoint hosts. In this case, ξ approximately equals one. Therefore, the average link available time can be simply expressed by $E(T_X)$, i.e.,

$$T_{HAV} \approx E(T_X) \tag{7}$$

2.3.2 Average Number of Link Changes for Maintaining of End-to-End Connection

It is clear that to provide an end-to-end connection in an ad hoc network, an established route path must be maintained until the communication task is completed. However, link outage in an ad hoc network becomes more significant compared to that in conventional mobile networks. Usually, any link outage on this existing path triggers a new route discover/recovery process, which may cause route change. On the other hand, in order to alleviate control overheads, shorten transmission delay and utilize the network transmission capacity efficiently, the link changes must be as little as possible. Therefore, it is important to study the average number of link changes for maintaining an end-to-end arbitrary communication pair in an ad hoc network

Let N_{pc} be the number of route changes to maintain an existing end-to-end connection in an ad hoc network and N_{hc} be the number of link changes experienced by a mobile host during an end-to-end connection. Since the mobile host in an ad hoc network is moving randomly, it is assumed that mobile link outage along the existing end-to-end route across the ad hoc network is independent. Therefore, the expected mean of N_{pc} can be approximated using

$$E(N_{pc}) \approx H_a E(N_{hc}), \tag{8}$$

where H_a denotes the average number of hops for an arbitrary end-to-end route which is given by equation (4), and $E(N_{hc})$ denotes the expected mean of N_{hc} can be obtained using the following calculation.

As shown in Figure 2-4, it is supposed that host h_j has an established link connection with host h_1 at time slot T_1. In order to keep an active link for h_j, when h_j leaves from the cell of h_1 at the beginning of time slot T_2, h_j needs to immediately handover the link connection to the following host h_2. If the cells of conjoint hosts h_i are overlapped each other, then h_j is able to continually switch its link at sojourn time T_i (i=1,2,$\cdots$), which follows the distribution of the sojourn time T_X. Let $f_{T_X}(t)$ be the PDF of T_X, then the Laplace transfers of $f_{T_X}(t)$ is given by

$$f^*_{T_X}(s) = \int_{-\infty}^{\infty} f_{T_X}(t) e^{-st} dt$$

Let random variable T_C denote the holding time of an end-to-end connection. If T_C is negative exponentially distributed, then the PDF of T_C is give by $f_{T_C}(t) = \mu_C e^{-\mu_C t}$ $(t > 0)$, where $1/\mu_C$ is the average connection holding time. Therefore,

$$f^*_{T_X}(\mu_C) = \int_0^{\infty} f_{T_X}(t) e^{-\mu_C t} dt = E(e^{-\mu_C t}) \tag{9}$$

The number of link changes experienced by h_j to keep an active link connection as it moves can be obtained by [63][65]:

$$E(N_{hc} \mid t_1, t_2, t_3 \cdots) = 1 \cdot \int_{t_1}^{t_1+t_2} \mu_C e^{-\mu_C t} dx + 2 \cdot \int_{t_1+t_2}^{t_1+t_2+t_3} \mu_C e^{-\mu_C t} dx + 3 \cdot \int_{t_1+t_2+t_3}^{t_1+t_2+t_3+t_4} \mu_C e^{-\mu_C t} dx + \cdots$$

$$= \int_{t_1}^{\infty} \mu_C e^{-\mu_C t} dx + \int_{t_1+t_2}^{\infty} \mu_C e^{-\mu_C t} dx + \int_{t_1+t_2+t_3}^{\infty} \mu_C e^{-\mu_C t} dx + \cdots$$

$$= e^{-\mu_C t_1} + e^{-\mu_C (t_1+t_2)} + e^{-\mu_C (t_1+t_2+t_3)} + \cdots$$

where the term $1 \cdot \int_{t_1}^{t_1+t_2} \mu_C e^{-\mu_C t} dx$ represents the first handover experienced by h_j during the time interval $[t_1, t_1+t_2]$, the term $2 \cdot \int_{t_1+t_2}^{t_1+t_2+t_3} \mu_C e^{-\mu_C t} dx$ represents the second handover during the time interval $[t_1+t_2, t_1+t_2+t_3]$, and so on.

Considering $E(N_{hc}) = E(E(N_{hc} \mid t_1, t_2, t_3 \cdots))$ and using equation (9), the expected mean value of N_{hc} can be given by

$$E(N_{hc}) = f^*_{T_X}(\mu_C) + f^{*\,2}_{T_X}(\mu_C) + f^{*\,3}_{T_X}(\mu_C) + \cdots = \frac{f^*_{T_X}(\mu_C)}{1 - f^*_{T_X}(\mu_C)} = \frac{1}{\frac{1}{f^*_{T_X}(\mu_C)} - 1} \tag{10}$$

In the rest of this section, the mobile handover model [1] and the random walk model [2][67] are used to derive the average link available time and the average number of link changes experienced by a mobile host during an end-to-end connection. Specially, the notation T_{X1} and N_{hc1} represent the mobility parameters derived by the mobile handover model. The notation T_{X2} and N_{hc2} denote the mobility parameters derived by the random walk model.

2.3.3 Mobility Analysis for the Mobile Handover Model

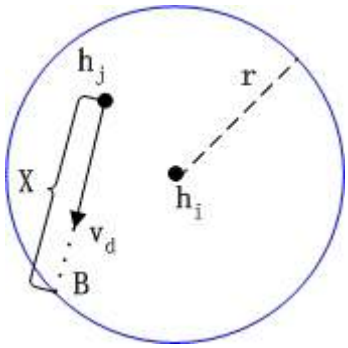


Figure 2-8. Host sojourn time for the mobile handover model

The first mobility model [1] for ad hoc networks is considered as that (1) the moving rate and the moving direction of mobile hosts are considered as independent random variables, and (2) mobile host has constant moving velocity while it is moving in the same cell, but the

moving velocity including both the moving rate and the moving direction are regenerated while the mobile host moves across boundaries from cell to cell. However, the moving rate of the mobile host is assumed to be generally distributed, and the moving direction of the mobile host is assumed to be uniformly distributed in the range of $(0, 2\pi)$.

As shown in Figure 2-8, the radio transmission range of h_i is r. A mobile host h_j moves in the cell of h_i starting from an arbitrary point A with a constant relative velocity. After traveling a distance of X in the cell, h_j leaves the cell at point B. Let v_d denote the relative moving rate of h_j with respect to h_i. The PDF of the distance X that h_j has traveled in the cell is given by [1], that is:

$$f_X(x) = \begin{cases} \frac{2}{\pi r^2}\sqrt{r^2 - (\frac{x}{2})^2}, 0 \le x \le 2r \\ 0, \quad \text{elsewhere} \end{cases} \tag{11}$$

Therefore, the sojourn time T_{X1} is given by $T_{X1} = X / v_d$. Its PDF can be obtained using equation (11), that is

$$f_{T_{X1}}(t) = \int_{-\infty}^{\infty} |v_d| f_X(tv_d) f_{v_d}(v_d) dv_d = \int_0^{\frac{2r}{t}} \frac{2v_d}{\pi r^2}\sqrt{r^2 - (\frac{tv_d}{2})^2} f_{v_d}(v_d) dv_d, \quad t \ge 0 \tag{12}$$

The cumulated CDF for T_{X1} can be given by

$$F_{T_{X1}}(t) = \int_0^t f_{T_{X1}}(m) dm$$

$$= 1 - \int_0^{\frac{2r}{t}} f_{v_d}(v_d)\left[1 - \frac{tv_d}{\pi r^2}\sqrt{r^2 - (\frac{tv_d}{2})^2} - \frac{2}{\pi} arcsin\frac{tv_d}{2r}\right] dv_d \tag{13}$$

In this case, the average link available time can be obtained as

$$E(T_{X1}) = \int_0^{\infty} t f_{T_{X1}}(t) dt = \int_0^{\infty}\int_0^{\frac{2r}{t}} \frac{2v_d t}{\pi r^2}\sqrt{r^2 - (\frac{tv_d}{2})^2} f_{V_d}(v_d) dv_d dt$$

$$= \frac{8r}{3\pi} E(1/v_d) \tag{14}$$

where $E(1/v_d)$ is the expected mean value of $1/v_d$.

Furthermore, let $f^*_{T_{X1}}(s)$ denote the Laplace transfers of $f_{T_{X1}}(t)$. The value of $f^*_{T_{X1}}(\mu_C)$ can be obtained as

$$f^*_{T_{X1}}(\mu_C) = \int_0^\infty f_{T_{X1}}(t) e^{-\mu_C t} dt = \frac{2}{\pi r^2} \int_0^\infty e^{-\mu_C t} \int_0^{\frac{2r}{t}} v_d f_{v_d}(v_d) \sqrt{r^2 - (\frac{tv_d}{2})^2} dv_d dt$$

$$= \frac{2}{\pi r^2} \int_0^\infty v_d f_{v_d}(v_d) \int_0^{\frac{2r}{v_d}} e^{-\mu_C t} \sqrt{r^2 - (\frac{tv_d}{2})^2} dt dv_d \tag{15}$$

According to equation (10), the average number of link changes experienced by h_j during an end-to-end connection is given by

$$E(N_{hc1}) = \frac{1}{\frac{1}{f^*_{T_{X1}}(\mu_C)} - 1} \tag{16}$$

2.3.4 Analysis for the Random Walk Model

The second model for mobility analysis in ad hoc networks is to consider that the moving velocity of mobile hosts including both of the moving rate and moving direction in a cell are randomly changed rather than that is constant as assumed in the analysis of the first model. This model is called random walk model [2], in which mobile host moves from an existing position to a new position in a random direction and at a random moving rate. Although the random walk model considers a more complicated situation but it matches the mobile host moving behavior in ad hoc networks.

In the following analysis, it is assumed that the movement epoch of a mobile host occurs at the each boundary of fixed time interval τ_e, so that the variation of mobile host moving velocity is a discrete function of time t. However, if the fixed time-interval is selected small enough, then the discrete moving velocity is able to match the mobile host real-time move. A new technical term called Relative Epoch (RE) is defined as the time interval, in which the relative movement between two mobile hosts h_i and h_j keeps a constant velocity vector $\mathbf{v}_d$. However, $\mathbf{v}_d$ can be changed randomly from RE to RE. Since h_i and h_j can change their movement asynchronously, in this case, the movement epoch change made by either h_i or h_j can trigger a new RE. The changes of RE is illustrated in Figure 2-9, where τ_a and $\tau_e - \tau_a$ are the durations of REs triggered by h_i and h_j, respectively. Since τ_e is small enough, in order to simplify the analysis, in the following analysis, it is assumed that the duration of RE uses the mean value as $\tau_d = \frac{\tau_e - \tau_a + \tau_a}{2} = \frac{\tau_e}{2}$. Therefore, RE can be expressed using a random mobility profile (v_d, θ_d, τ_d), where $v_d = |\mathbf{v}_d|$ is the relative moving rate between h_i and h_j, θ_d is the relative moving direction between h_i and h_j, and $\tau_d = \tau_e / 2$ is the average time interval of RE.

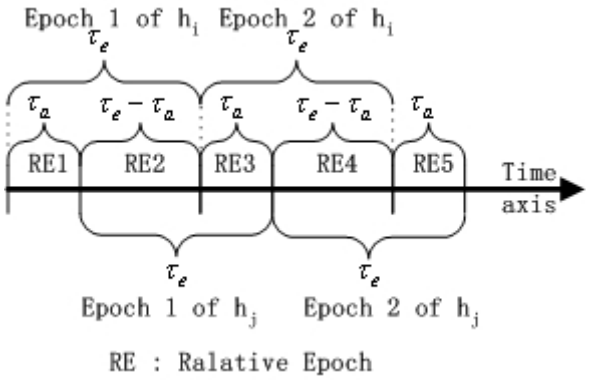


Figure 2-9. The relative epoch of h_i and h_j

As shown in Figure 2-10, mobile host h_j starts random walks in the cell of the mobile host h_i from an arbitrary point A for $n-1$ REs (n=1,2,...), and walks out of the cell coverage range at point B in the n th RE. In this case, the sojourn time T_{X2} is equal to the time that h_j remains inside of the cell of h_i.

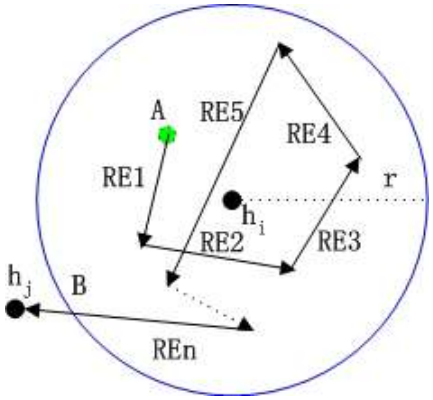


Figure 2-10. Host sojourn time for random walk model

Let $\Pr_{X'n}$ (n=1,2,...) denote the probability that h_j leaves the cell of h_i in the n th RE. $\Pr_{X'n}$ equals the probability that h_j resides in the cell for the first $n-1$ REs and leaves the cell during the n th RE. Since h_j moves randomly in the cell of h_i, the event that h_j leaves this cell during individual RE is independent, $\Pr_{X'n}$ can be given by:

$$\Pr_{X'n} = \Pr_{X'}{}^{n-1}(1-\Pr_{X'}), \quad (n=1,2,...) \tag{17}$$

where $\Pr_{X'}$ denotes the probability that h_j resides in the cell of h_i during a single RE. Since the random walk of h_j in the cell of h_i during a single RE is equivalent to the movement of h_j in this cell by using the first mobile handover model, the probability that h_j leaves the cell of h_i during a single RE equals the probability that the sojourn time T_{X1} is smaller than the duration of one RE, i.e.,

$$1-\Pr_{X'} = \Pr(T_{X1} < \tau_d)$$

Thus, $\Pr_{X'} = 1 - \Pr(T_{X1} < \tau_d) = 1 - F_{T_{X1}}(\frac{\tau_e}{2})$,

where $F_{T_{X1}}(t)$ is the CDF of the host sojourn time T_{X1} for the first model. Using equation (13), $\Pr_{X'}$ can be given by

$$\Pr_{X'} = \int_0^{\frac{4r}{\tau_e}} f_{V_d}(v_d)\left[1 - \frac{\tau_e v_d}{2\pi r^2}\sqrt{r^2 - (\frac{\tau_e v_d}{4})^2} - \frac{2}{\pi} arcsin\frac{\tau_e v_d}{4r}\right] dv_d \qquad (18)$$

Let $T_{X'n}$ $(n=1,2,...)$ be the sojourn time that h_j resides in the cell of h_i until h_j leaves in the n th RE. Obviously, $T_{X'n}$ consists of the duration of n-1 REs and the time interval that h_j resides inside of the cell of h_i during the n th RE, which is denoted as T_Y. Consequently,

$T_{X'n} = (n-1)\tau_d + T_Y$, $(n=1,2,...)$

The PDF of $T_{X'n}$ can be obtained as

$f_{T_{X'n}}(t) = f_{T_Y}[t-(n-1)\tau_d]$

where $f_{T_Y}(t)$ is the PDF of T_Y.

The average value of $T_{X'n}$ can be derived as

$$E(T_{X'n}) = (n-1)\tau_d + E(T_Y) = \frac{(n-1)\tau_e}{2} + E(T_Y), \quad (n=1,2,...) \qquad (19)$$

On the other hand, the moving that h_j leaves the cell in last RE is similar to the moving that h_j travels in the cell of h_i by using the first mobile handover mode. Considering $T_{X'} \le \tau_d$, T_Y can be expressed by T_{X1}, that is:

$$T_Y = \begin{cases} T_{X1}, 0 \le t \le \tau_d \\ 0, \text{ otherwise} \end{cases}$$

Then, the PDF of T_Y can be given by

$$f_{T_Y}(t)=\begin{cases}\dfrac{f_{T_{X1}}(t)}{F_{T_{X1}}(\tau_d)}, 0\le t\le \tau_d\\ 0, \quad \text{otherwise}\end{cases},$$

The expected mean value of T_Y is obtained using equation (12), that is

$$E(T_Y)=\int_0^{\infty} t f_{T_Y}(t)dt=\frac{1}{F_{T_{X1}}(\frac{\tau_e}{2})}\int_0^{\frac{\tau_e}{2}}\int_0^{\frac{2r}{t}}\frac{2v_d t}{\pi r^2}\sqrt{r^2-(\frac{tv_d}{2})^2}\,f_{V_d}(v_d)dv_d dt$$

$$=\frac{8}{3\pi F_{T_{X1}}(\frac{\tau_e}{2})}\left\{rE(\frac{1}{v_d})-\frac{1}{r^2}\int_0^{\frac{4r}{\tau_e}}\frac{f_{V_d}(v_d)}{v_d}\left[r^2-(\frac{\tau_e v_d}{4})^2\right]^{\frac{3}{2}}dv_d\right\} \qquad (20)$$

Based on the above analysis, the PDF of the sojourn time T_{X2} is expressed as

$$f_{T_{X2}}(t)=\sum_{n=1}^{\infty}\Pr_{X'n}f_{T_{X'n}}(t)=\sum_{n=1}^{\infty}\Pr_{X'}{}^{n-1}(1-\Pr_{X'})f_{T_{X'}}[t-\frac{(n-1)\tau_e}{2}] \qquad (21)$$

The average link available time can be obtained as:

$$E(T_{X2})=\sum_{n=1}^{\infty}\Pr_{X'n}E(T_{X'n})$$

Substituting (17) and (19), $E(T_{X2})$ can be given by

$$E(T_{X2})=\sum_{n=1}^{\infty}\Pr_{X'}{}^{n-1}(1-\Pr_{X'})\left[\frac{(n-1)\tau_e}{2}+E(T_Y)\right]$$

$$=\frac{\tau_e}{2}\sum_{n=1}^{\infty}(\Pr_{X'}{}^{n-1}-\Pr_{X'}{}^{n})(n-1)+E(T_Y)\sum_{n=1}^{\infty}(\Pr_{X'}{}^{n-1}-\Pr_{X'}{}^{n})$$

$$=\frac{\tau_e\Pr_{X'}}{2(1-\Pr_{X'})}+E(T_Y), \qquad (22)$$

where $\Pr_{X'}$ and $E(T_Y)$ can be obtained by equation (18) and (20), respectively.

Moreover, let $f^*_{T_{X2}}(s)$ represent the Laplace transfers of $f_{T_{X2}}(t)$. $f^*_{T_{X2}}(\mu_C)$ can be obtained using equation (21), that is:

$$f^*_{T_{X2}}(\mu_C) = \int_0^\infty f_{T_{X2}}(t)e^{-\mu_C t}dt = \int_0^\infty \left\{ \sum_{n=1}^\infty \Pr_{X'}{}^{n-1}(1-\Pr_{X'}) f_{T_Y}[\text{t-}(n\text{-}1)\tau_d] \right\} e^{-\mu_C t}dt$$

$$= \sum_{n=1}^\infty \Pr_{X'}{}^{n-1}(1-\Pr_{X'}) \int_{-\infty}^\infty f_{T_Y}[\text{t-}(n\text{-}1)\tau_d] e^{-\mu_C t}dt$$

$$= (1-\Pr_{X'}) f^*_{T_Y}(\mu_C) \sum_{m=0}^\infty \Pr_{X'}{}^m e^{-\mu_C \tau_d m}$$

Since $0 < P_{X'} < 1$ and $e^{-\mu_C \tau_d} < 1$, we have

$$\sum_{m=0}^\infty \Pr_{X'}{}^m e^{-\mu_C \tau_d m} = \frac{1}{1-\Pr_{X'} e^{-\mu_C \tau_d}}$$

Consequently, $f^*_{T_{X2}}(\mu_C) = \frac{1-\Pr_{X'}}{1-\Pr_{X'} e^{-\mu_C \tau_d}} f^*_{T_Y}(\mu_C)$, (23)

where the term $f^*_{T_Y}(\mu_C)$ represents the value of the corresponding Laplace transfers of $f_{T_Y}(t)$ with a particular value that $s = \mu_C$. It can be given by:

$$f^*_{T_Y}(\mu_C) = \int_0^\infty f_{T_Y}(t)e^{-\mu_C t}dt$$

$$= \frac{2}{\pi r^2 F_{T_{X1}}(\frac{\tau_e}{2})} \int_0^{\frac{\tau_e}{2}} \int_0^{\frac{2r}{t}} v_d e^{-\mu_C t} \sqrt{r^2 - (\frac{t v_d}{2})^2} f_{V_d}(v_d) dv_d dt$$

According to equation (10) and (23), the average number of link changes experienced by h_j during an end-to-end connection is given by

$$E(N_{hc2}) = \frac{1}{\frac{1}{f^*_{T_{X2}}(\mu_C)} - 1} = \frac{1}{\frac{1-\Pr_{X'} e^{-\mu_C \tau_d}}{(1-\Pr_{X'}) f^*_{T_Y}(\mu_C)} - 1}$$ (24)

It is obvious that the formulas derived by the random walk model are more complicated than those by the first handover model. The next section provides the numerical results obtained by the two mobility models.

2.4 Evaluation and Results

In most simulation experiments [2][26][27] for the investigation of ad hoc networks, the host random moving is assumed that during each moving epoch, the moving rates (v_i and v_j) and directions (θ_i and θ_j) of mobile hosts h_i and h_j are uniformly distributed among the ranges of $[0, V_m]$ and $[0, 2\pi]$, respectively. In this case, the moving rate of the relative velocity $\mathbf{v}_d$ between two adjacent mobile hosts h_i and h_j is given by

$$v_d = |\mathbf{v}_d| = |\mathbf{v}_j - \mathbf{v}_i| = \sqrt{v_i^2 + v_j^2 - 2v_i v_i \cos(\theta_i - \theta_j)}$$

The PDF of v_d is derived in Appendix B, that is:

$$f_{V_d}(v_d) = \begin{cases} \dfrac{2v_d}{\pi V_m^2} \displaystyle\iint_{\substack{0 \le x \le V_m \\ 0 \le y \le V_m \\ |x-y| < v_d < x+y}} \dfrac{1}{\sqrt{4x^2y^2 - (v_d^2 - x^2 - y^2)^2}} dxdy, & 0 \le v_d \le 2V_m \\ 0, & \text{elsewhere} \end{cases} \tag{25}$$

The average relative moving rate is $E(v_d) \approx 0.724 V_m$, and $E(1/v_d) \approx 2.31/V_m$.

Figure 2-11 illustrates the PDF of the relative moving rate v_d. It can be found that the analysis results matches simulation results (10^6 times trials) very well. The CDF of v_d is shown in Figure 2-12. It can be seen that the relative moving rate follows neither uniform distribution nor Rayleigh distribution, however, these two distributions have been used in [66][76] for the performance analysis.

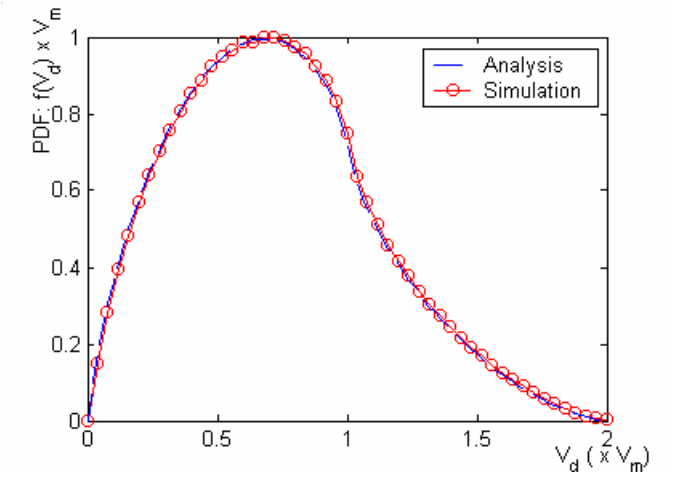


Figure 2-11. The PDF of v_d

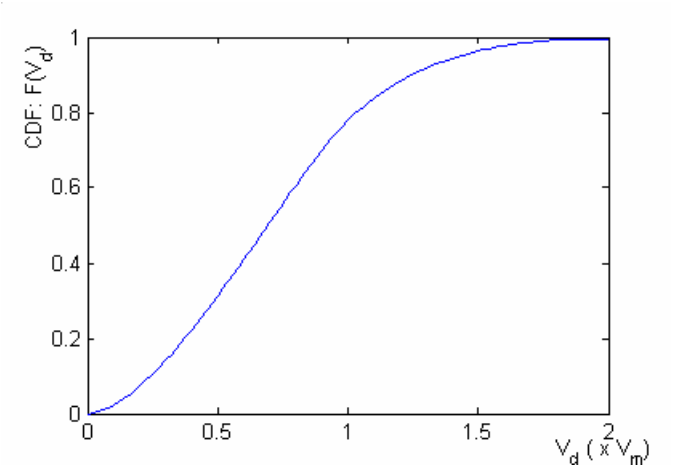


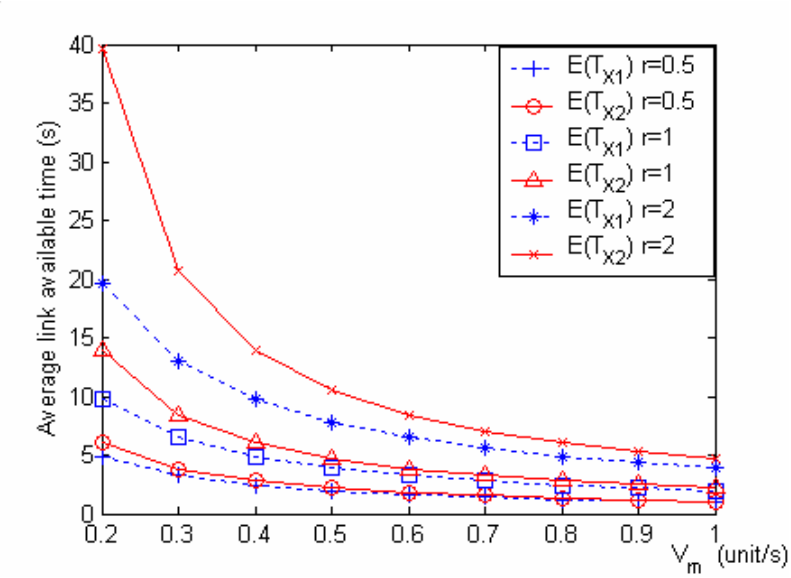
Figure 2-12. The CDF of v_d

2.4.1 The Average Link Available Time

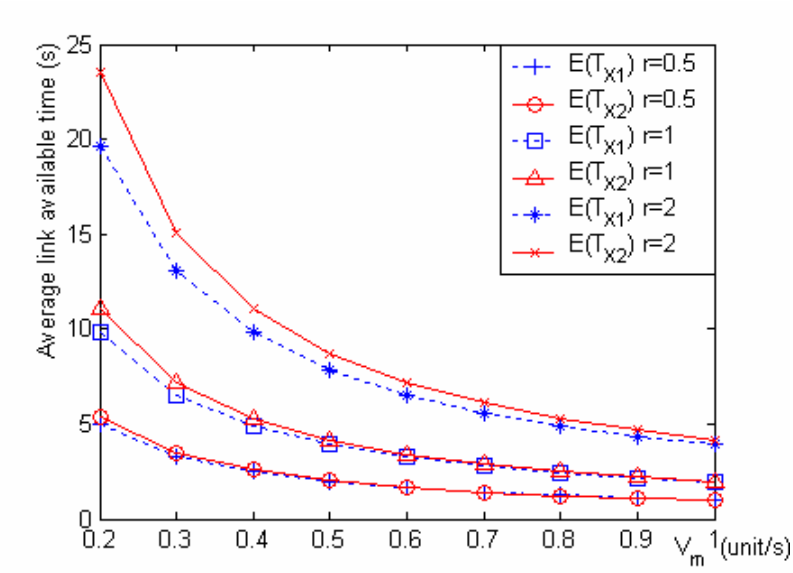
Figure 2-13 (a) and (b) show the numerical values of the average link available time versus maximum moving rate V_m, with different transmission range r. In this figure, $E(T_{X1})$ is obtained from the first mobile handover model using equation (14), and $E(T_{X2})$ are derived from the random walk model using equation (22). It can be seen for a fixed value of the maximum host-moving rate V_m, link available time increases with r. This is because that

the increase of transmission range leads to the increase of distance traveled by neighboring hosts before wireless link is outage. On the other hand, link available time decreases with the increase of the maximum host-moving rate V_m. $E(T_{X2})$ is always larger than $E(T_{X1})$ when r and V_m have fixed values. This is because in random walk model, the relative velocity between two neighboring hosts can be changed while a host randomly walks inside of a cell. Compared to the moving at constant velocity in the mobile handover model, the random walk at different moving velocity from RE to RE causes that mobile hosts can reside longer in a cell.

Figure 2-13 (a) and (b) also show the difference between $E(T_{X1})$ and $E(T_{X2})$ for $\tau_e = 1s$ and $\tau_e = 0.2s$, respectively. When r or V_m increases, the difference between $E(T_{X1})$ and $E(T_{X2})$ increases correspondingly. On the other hand, from the comparison of Figure 2-13 (a) and (b), it can be seen that the value of $E(T_{X2})$ increases when the duration of RE decreases. If the value of τ_e is small, then the host must frequently change the moving velocity. In this case, the mobile host has more chances to keep itself inside of the cell comparing to the situation of large τ_e. From Figure 2-13 (a), it can be seen that when $\tau_e = 0.2s$, there is large difference between $E(T_{X1})$ and $E(T_{X2})$, especially when r has a large value. For example, when $r = 2$ and $V_m = 0.2$ unit/s, $E(T_{X2})$ is twice of $E(T_{X1})$. However, as shown in Figure 2-13 (b), when $\tau_e = 1s$ and $V_m \geq 0.5$ unit/s, it is difficult to distinguish the values of $E(T_{X1})$ and $E(T_{X2})$. Therefore, it can be concluded that in an ad hoc network, when mobile hosts move at fast moving rates, or hosts change their velocity slowly, the performances evaluated by using the mobile handover model and the random walk model are equivalent.



(a) $\tau_e = 0.2s$



(b) $\tau_e = 1s$

Figure 2-13. Numerical results of average link available time

A discrete-event simulator using C++ language is developed to validate the numerical results derived by the two mobility models. Since the random walk model can flexibly simulate diverse host movement by changing the duration of moving epochs, this model is used in the simulation experiments. Totally $N = 200$ mobile hosts with the transmission range $r = 1$

unit are uniformly located in a square area with $S = 10 \times 10$ square units. The maximal moving rate V_m of mobile hosts is ranging from 0.2 to 1 unit/s, and the duration of each movement epoch varies from 0.1 to 1s. Moreover, the boundless simulation area [2] is applied to avoid the border effect [62]. Hosts that reach one side of the network continue traveling and reappear on the opposite side. Thus, hosts are able to travel unobstructedly in the network.

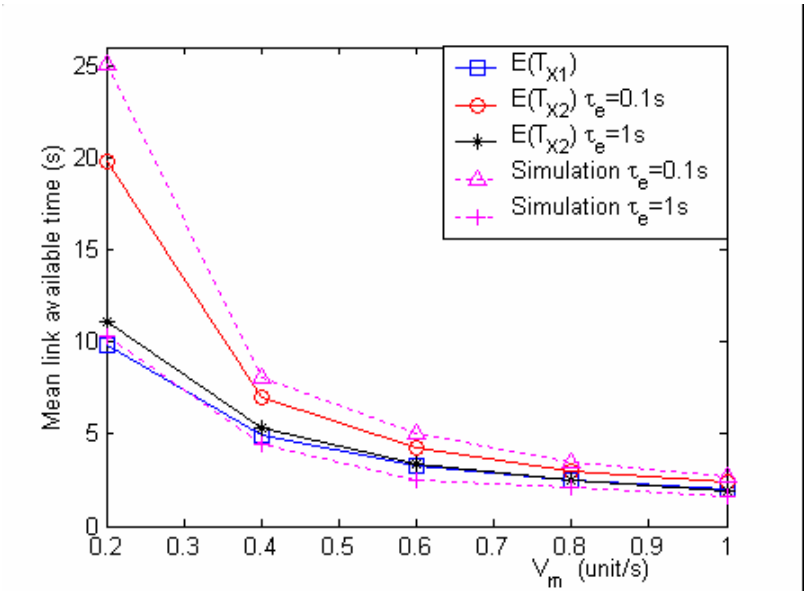


Figure 2-14. Average link available time

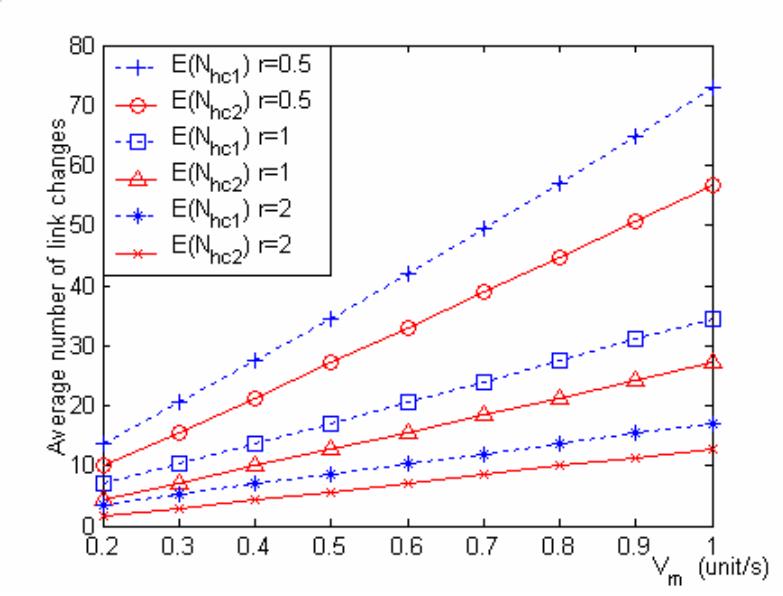
$N = 200$, $l = 10$ units, $r = 1$ unit

Figure 2-14 shows both the analytical values and simulation results of the average link available time, for $\tau_e = 0.1s$ and $\tau_e = 1s$, respectively. It can be seen that when $\tau_e = 0.1s$, both $E(T_{X1})$ and $E(T_{X2})$ are very close to the simulation results. However, when $\tau_e = 1s$, $E(T_{X2})$ can always match the simulation results, but $E(T_{X1})$ can approximate the simulation values only when $V_m \geq 0.8$. Therefore, using the random walk model, $E(T_{X2})$ can correctly formulate the link available time for different moving scenarios. By contrast,

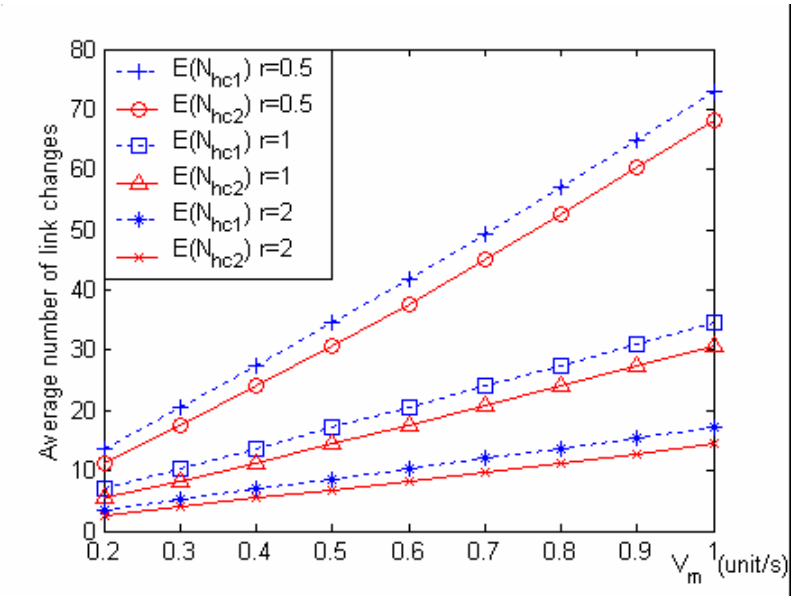
the mobile handover model can be used to characterize host movement only if hosts move at relatively higher speeds with less variation.

2.4.2 The Average Number of Link Outage

The average number of link changes experienced by one mobile host during each end-to-end connection is evaluated under the assumption that the average holding time of a connection is $1/\mu_C = 60s$. Figure 2-15 shows the numerical results of $E(N_{hc1})$ and $E(N_{hc2})$, which are derived by the mobile handover model and the random walk model, respectively. It is clear that when V_m increases, the average number of link changes increases following approximately linear rate. This is because when host moving rate becomes fast, mobile hosts change their wireless links frequently. Fortunately, the increase of the radio transmission range r is able to reduce the link change rate. As shown in Figure 2-15, both $E(N_{hc1})$ and $E(N_{hc2})$ largely decrease while r increases. For example, the values of $E(N_{hc1})$ for $r = 2$ units are about twice of those for $r = 1$ unit.



(a) $\tau_e = 0.2s$



(b) $\tau_e = 1s$

Figure 2-15. Numerical results of $E(N_{hc1})$ and $E(N_{hc2})$

Figure 2-15 (a) and (b) illustrate the difference between $E(N_{hc1})$ and $E(N_{hc2})$, for $\tau_e = 1s$ and $\tau_e = 0.2s$, respectively. When V_m has a fixed value, $E(N_{hc2})$ is always smaller than $E(N_{hc1})$, since host sojourn time in a cell derived by the random walk mobility model is longer than that derived by the handover mobility model. Moreover, the difference between $E(N_{hc1})$ and $E(N_{hc2})$ for $\tau_e = 0.2s$ is much larger than that for $\tau_e = 1s$. This is because that in the random walk mobility model, the host sojourn time increases when τ_e decreases.

Figure 2-16 shows both simulation results and the numerical results of $E(N_{hc1})$ and $E(N_{hc2})$ for $\tau_e = 0.1s$ and $\tau_e = 1s$. It can be found that both the simulation results and the numerical results increase following a linear rate while V_m increases. There is large difference between $E(N_{hc1})$ and the values of simulation results, and the difference increases with the decrease of τ_e. Contrarily, $E(N_{hc2})$ matches the relevant simulation

results well. On the other hand, as shown in Figure 2-16, most values of the simulation results are lower than the numerical results of $E(N_{hc2})$. This is because during the simulation, some mobile hosts may lose the wireless links to connect their neighboring hosts temporarily. In this case, the actual holding times of end-to-end connections are reduced, and the statistic results of link changes should be smaller than the expected values.

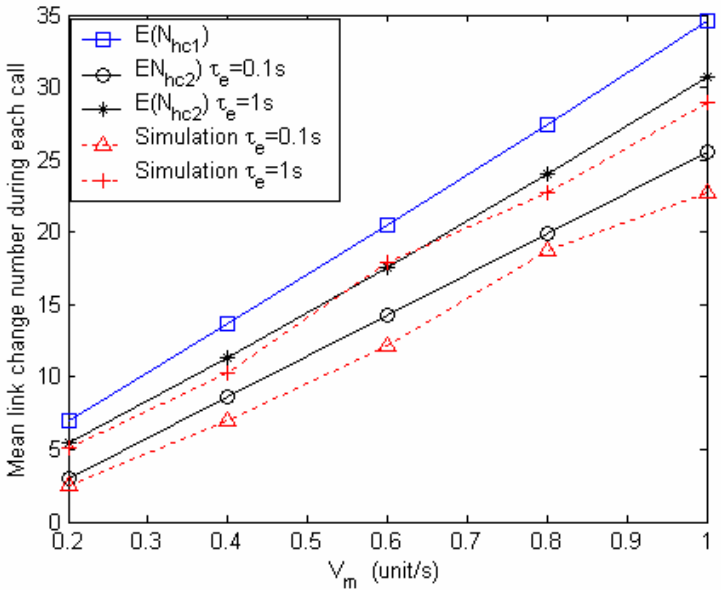


Figure 2-16. The average number of link changes

$N = 200$, $l = 10$ units, $r = 1$ unit, $1/\mu_C = 60s$

2.5 Conclusion

This chapter presents analytical derivations of some statistic characteristics of a homogenous ad hoc network with respect to: (1) the average number of neighboring hosts of each host, (2) the average number of hops for an arbitrary communication pair, (3) the link available time and (4) the average number of path/link changes for each connection. Especially, the latter two mobility related parameters are derived using mobile handover model and the random walk model. The analytical results can be applied to design suitable transmission range,

assign wireless channel, set proper thresholds in routing protocols, arrange location update in location service protocols and validate simulation scenario.

In order to convert the two-body mobility problem in ad hoc networks to one-body problem, the relative velocity between two adjacent hosts is employed to investigate host mobility. This study focuses on the performance analysis of ad hoc mobility that hosts are randomly moving at speeds with general distribution as well as a specific moving case where the host random moving including speed and direction is assumed to be uniform distribution. The simulation results show good agreement with the analytical values. Although the random walk analytical model is more complicated, it can better describe the complicated host movement in an ad hoc network. The analysis based on random walk model can better evaluate the performance of network than that based on the mobile handover model. On the other hand, the mobile handover model can be simply applied to characterize host mobility when hosts move at relatively higher speeds with less variation. Although the analysis in this chapter only aims at the mobile handover model and the random walk model, the derivation methods can be used to formulate the characteristics of other mobility models for ad hoc networks.

Chapter 3 k-hop Compound Metric Based Clustering Scheme

3.1 Introduction

The multihop connection across an ad hoc network needs to be often rerouted due to the random walks of the mobile hosts. In this case, the flat routing protocols such as Dynamic Source Routing (DSR) [44] and Ad hoc On-demand Distance Vector Routing (AODV) [45]) may not be able to handle a large scale ad hoc network, especially when the density of mobile hosts in the network is heavy. By contrast, hierarchical routing using cluster structure, grid based routing protocol [32] or Zone Routing Protocol (ZRP) [38] can be used to solve such problem for large scale ad hoc networks.

The most popular way of building hierarchy is to group hosts geographically close to each other into clusters. Compared with other strategies for grouping mobile hosts by using fixed geographic scopes such as grid [32] or circle [77], clustering is more flexible to build hierarchical network topology. The goal of clustering network architecture is to reduce the overhead being used for providing network topology details so that the network becomes scalable. The clustering architecture is able to provide many important management procedures such as channel access [39][78], power control [79] and mobility management [41][66][80]. This chapter focuses on the design of a novel approach of dynamic k-hop clustering architecture, which is called k-hop Compound Metric Based Clustering (KCMBC). This approach is able to demonstrate significant improvement on the network performance in terms of scalability and stability for large scale ad hoc networks.

3.2 A Survey on Related Works

In multihop ad hoc networks, the aggregation of hosts into clusters provides a convenient framework for the development of important feature such as routing, bandwidth allocation, mobility management and location management. In clustering network architecture, cluster-heads are elected among the mobile hosts to be the local coordinators. Each cluster-head needs to maintain the membership of its cluster and track the local topology changes due to host mobility. Frequent cluster-head changes adversely affect the network performance and result in a lot of control overheads. Moreover, the election of cluster-head may require large network convergence time. Hence, an efficient clustering scheme to keep the topology to be stable becomes even more important.

There are two kinds of distributed clustering algorithms. The first one is called the lowest ID algorithm [39][40][78] in which the host with the lowest network ID in the network cluster is elected as the cluster-head. The second one is called the highest connectivity (degree) algorithm [39][40], in which the host has the most number of neighboring hosts connected to it becomes the cluster-head. Comparing these two algorithms, the second approach can result in a high turnover of cluster-leader that is undesirable due to the high overhead associated with the change of cluster-leader. Cluster-leader-Gateway Switch Routing (CGSR) [40] modifies the highest connectivity algorithm to keep clusters unchanging as long as possible. Accordingly, the communication overhead is largely reduced. However, cluster-heads have to be involved in determining routing path and forwarding every packet, which create heavy load to cluster-heads. Hence CGSR is not scalable for a large diameter network.

Hierarchical State Routing (HSR) [41] is a multilevel clustering based link state routing protocol. It maintains a logical hierarchical topology by using the clustering scheme recursively. Hosts at the same logical level are grouped into clusters. The elected cluster-heads at the lower level become members of the next higher level. Although HSR has good scalability, the complicated management strategy makes it hard to implement for practical use.

Lin and Gerla [78] described a better lowest ID algorithm that creates non-overlapping clusters. Each host in the network broadcasts its clustering decision exactly once during a cluster construction process. Cluster-heads only take charge of clustering formation and update. Hence the bottleneck problem of cluster-heads is solved. In [81], the clustering algorithm was extended to a Connectivity Based k-hop Clustering (CBKC), where all the hosts of a cluster within distance at most k hops from the cluster-head. It is obvious that the cluster size (the number cluster member in the cluster) of k-hop cluster is more than that of 1-hop cluster. Fewer clusters are constructed by using k-hop clustering algorithm. Thus, compared with 1-hop clustering, k-hop clustering has more scalability for a large scale ad hoc network.

Another k-hop cluster formation algorithm is the Max-Min heuristic, which is proposed by Amis and Cetra [55]. This heuristic converges very fast with only $2k$ rounds of packets exchanged by each host. However, it ignores the effect of host mobility during cluster formation. The clusters may be reconstructed frequently when hosts in the same cluster move towards different directions. Hence the Max-Min heuristic cannot create good cluster structure in a dynamic environment. Moreover, since the converge-cast applied in the Max-Min heuristic may lose the information of hosts, which are located on the cluster border,

these hosts cannot join their selected cluster correctly (the detail can be found in Section 3.3.3).

Mobility of hosts causes clusters to get disrupted and thus triggers reclustering. Therefore, the use of mobility information for cluster formation is a reasonable proposition. McDonald [66] bounded cluster formation with the probability of path availability. A mobility model was developed and used to derive expressions for the probability of path availability as a function of time. However, it may be hard to implement such a complex algorithm in a real ad hoc network. Moreover, the Mobility-Based Clustering (MBC) approach [80] used a combination of both physical and logical partitions of the network, i.e. geographic proximity and functional relation between hosts, such as mobility pattern. The performance evaluation via simulation demonstrated that MBC scheme improves the stability of the cluster algorithm.

None of the above clustering schemes shows good performance on both scalability and stability. To generate robust cluster structure for a dynamic ad hoc network with large scale, a novel k -hop Compound Metric Based Clustering approach is proposed in this chapter.

3.3 The KCMBC Approach

A k -hop cluster C_m is defined as a set of hosts under the same cluster-head h_m. The maximum number of hops from a mobile host in the cluster to the cluster-head is k. Using the KCMBC approach, an ad hoc network can be divided into non-overlapping clusters. As shown in Figure 3-1, each mobile host either belongs to a unique cluster, or is an orphan host which cannot join a cluster temporarily. Every cluster consists of one cluster-head, ordinary

cluster members which are located inside of a cluster, and gateways which are located on the cluster border and connect with neighboring clusters.

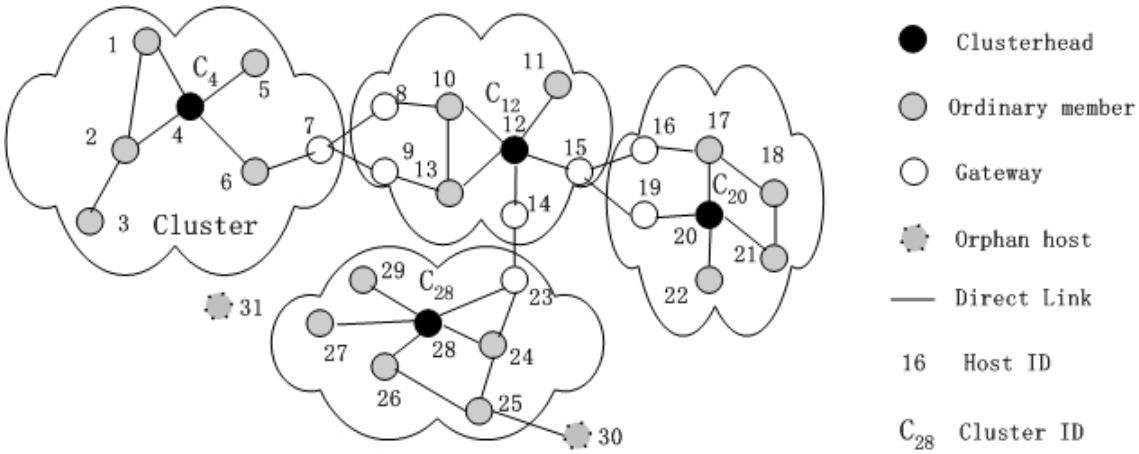


Figure 3-1. k -hop clusters created by KCMBC, $k = 2$

The proposed KCMBC applies the similar cluster formation procedure of the Max-Min k-Cluster heuristic [55]. In Max-Min heuristic, a mobile host h_i becomes a cluster-head if at least one of the following conditions is satisfied: (1) it has the highest host ID in k-hop neighborhood; (2) h_i does not have the highest host ID in its k-hop neighborhood, but there is at least one neighboring host h_j such that h_i has the highest host ID in h_j's k-hop neighborhood. The heuristic has four logical stages: In the first stage, the largest host ID in each host's k-hop neighborhood is propagated using k rounds of single-hop broadcasts; In the second stage, smaller IDs within the k-hop neighborhood are propagated via k rounds of single-hop broadcast; The third stage determines cluster-heads according to the above criteria; In the fourth stage, gateways initiate converge-cast which is used to deliver the cluster member information to the cluster-head.

As Max-Min heuristic has fast convergence with $O(k)$ rounds time complexity, it has good scalability for large diameter networks. However, the lowest ID algorithm applied in Max-

Min heuristic does not take into account the connectivity of hosts, and may produce more clusters than necessary. Moreover, due to lacking of mobility consideration, the Max-Min heuristic cannot provide satisfied performance in a dynamic environment. By contrast, the proposed KCMBC combines both the highest connectivity and host mobility to elect cluster-heads. Differing from other mobility-based clustering schemes [66][80], the KCMBC approach applies the average link expiration time to be mobility metric.

3.3.1 Mobility Metric

Because of the free movement of mobile hosts, link outage may result in cluster re-affiliation, and cause a lot of communication overhead. It is obvious that a mobile host, which has a high relative moving rate relevant to its neighboring hosts, is unsuitable to act as cluster-head. Therefore, a robust clustering scheme should utilize the relative moving information with respect to its neighboring hosts. In the KCMBC approach, the average link expiration time is used to predict the stability of the links between a host and its neighboring hosts.

The KCMBC approach is based on host mobility and the availability of position information through low power low cost GPS receivers or some other type of position service [33][34]. Host h_i broadcasts its current position to the neighboring hosts through periodical beacons. Let $\mathbf{s}_i(t)$ denote the position vector of h_i at time t, and τ_s is the time interval between two successive beacons. Accordingly, the neighboring hosts calculate h_i's velocity vector which is given by

$$\mathbf{v}_i(t) = \frac{\mathbf{s}_i(t) - \mathbf{s}_i(t-\tau_s)}{\tau_s},$$

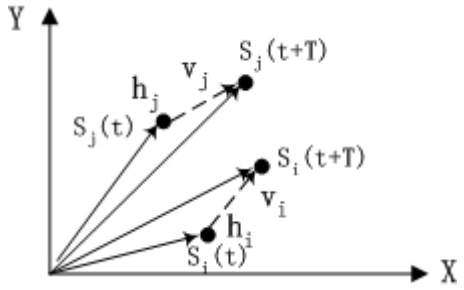


Figure 3-2 Relative position between h_i and h_j

Figure 3-2 shows two neighboring hosts h_i and h_j, which have the same radio transmission range of radius r. It is assumed that the distance between h_i and h_j is $|\mathbf{s}_i(t)-\mathbf{s}_j(t)|<r$ at time t. Let $\mathbf{v}_i(t)$ and $\mathbf{v}_j(t)$ denote the velocity vectors of h_i and h_j at time t, respectively. Then, at time $t+T$, the two hosts move to the new positions which can be expressed by

$$\begin{cases}\mathbf{s}_i(t+T)=\mathbf{s}_i(t)+T\mathbf{v}_i(t)\\ \mathbf{s}_j(t+T)=\mathbf{s}_j(t)+T\mathbf{v}_j(t)\end{cases}.$$

Let T_{ij} denote the link expiration time of link (h_i,h_j), then T_{ij} can be derived when the relative distance between h_i and h_j equals r, i.e.:

$$|\mathbf{s}_i(t+T_{ij})-\mathbf{s}_j(t+T_{ij})|=|\mathbf{s}_i(t)-\mathbf{s}_j(t)+T_{ij}[\mathbf{v}_i(t)-\mathbf{v}_j(t)]|=r \quad (26)$$

Using planar coordinates to express a vector, then

$\mathbf{s}_i\equiv(s_{ix},s_{iy})$ and $\mathbf{v}_i(t)\equiv(v_{ix},v_{iy})$,

where s_{ix} and v_{ix} are the components in the x direction, and s_{iy} and v_{iy} are the components in the y direction. The link expiration time can be obtained by deriving the solution of equation (26)

$$T_{ij}=\begin{cases}T_1, \text{ if } T_1>0\\ T_2, \text{if } T_2>0\end{cases} \quad (27)$$

where $T_{1,2} = \dfrac{\pm\sqrt{r^2(\Delta_{vx}^2 + \Delta_{vy}^2) - (\Delta_{sx}\Delta_{vy} - \Delta_{sy}\Delta_{vx})^2} - \Delta_{sx}\Delta_{vx} - \Delta_{sy}\Delta_{vy}}{\Delta_{vx}^2 + \Delta_{vy}^2}$,

and $\Delta_{sx} = s_{ix} - s_{jx}$, $\Delta_{sy} = s_{iy} - s_{jy}$, $\Delta_{vx} = v_{ix} - v_{jx}$ and $\Delta_{vy} = v_{iy} - v_{jy}$

Moreover, let T_i denote the average expiration time of all the links connected to h_i, then

$$T_i = \frac{1}{d_i} \sum_{j \in N(h_i)} T_{ij} , \qquad (28)$$

where d_i is the host degree (the number of neighboring hosts of h_i), and $N(h_i)$ denotes the set of h_i's neighboring hosts.

It is obvious that the host with a large value of average link expiration time is able to keep long connection with their neighboring hosts. Such a host should have more opportunity to be cluster-heads than the host with small average link expiration time. Therefore, the average link expiration time of each host, which can be easily calculated in a distributed manner, is used as mobility metric for cluster-head election.

3.3.2 Cluster-Head Election

In order to improve the stability in the clustered topology, the cluster-heads must be selected carefully. Considering a k-hop cluster, if the cluster-head has a large number of 1-hop neighboring hosts, and the links between the cluster-head and its neighboring hosts can keep available for a long period of time, then the probability of changing cluster-head becomes low. In this case, the cluster structure becomes stable regardless there are other mobile hosts joining or leaving this cluster. Therefore, the stability of the direct links connecting with the cluster-head is more important than that of other intra-cluster links.

In the proposed KCMBC scheme, firstly, the average link expiration time of each host is used to select cluster-head candidates. Host h_i is elected as a cluster-head candidate, if and only if the average link expiration time $T_i \geq T_{ALT}$, where T_{ALT} is a predefined mobility threshold. The average link available time $E(T_X)$ (see Chapter 2) can be used to set the suboptimal value of T_{ALT}.

If host h_i is elected as a cluster-head candidate, it then uses a compound metric $CP_i = (d_i, T_i, i)$ to compete for the role of cluster-head within its k-hop neighborhood, where the host degree d_i is the top priority parameter to reduce the opportunity to generate small clusters, the average link expiration time T_i is the second priority parameter, and the last parameter i is the host ID. Consequently, for any two candidates h_i and h_j, the term of $CP_i > CP_j$ can be expressed as

$$\{CP_i > CP_j\} \equiv \{(b_i > b_j) \cup ((b_i = b_j) \cap (T_i > T_j)) \cup ((b_i = b_j) \cap (T_i = T_j) \cap (i > j))\},$$

where the term $b_i > b_j$ represents that the number of h_i'neighboring hosts is more than the number of h_j'neighboring hosts, the term $(b_i = b_j) \cap (T_i > T_j)$ represents that h_i and h_j have the same number of neighboring hosts but the average link expiration time T_i is larger than T_j, and the term $(b_i = b_j) \cap (T_i = T_j) \cap (i > j)$ represents h_i and h_j have the same number of neighboring hosts and the same average link expiration time, but h_i's host ID is higher than h_j's ID.

Based on the compound metric CP, the cluster election procedure can be described as follows.

I. Initially, each cluster-head candidate sets its winning metric equal to its own compound metric, and the other non-cluster-head candidates set their winning metric equal to $CP = (0,0,0)$.

II. Every host locally broadcasts the winning metric to its neighboring hosts. When the host receives the winning metrics from the neighboring hosts, it upgrades the winning metric as the largest winning metric that it has received. In this case, for a k-hop cluster, the largest metric will be propagated k hops in order to ensure that all hosts have the same maximum winning metric for cluster. This process is denoted as Flood-max..

III. After k rounds of Floodmax, a host takes k rounds of single-hop broadcasts (Floodmin) to propagate the winning metrics. The Floodmin allows that the candidates with smaller metrics have the chance to construct their clusters. Again, each host records the new winning metric after each broadcast round, which is the smallest value among its own winning metric and the winning metrics received in this round.

IV. The cluster-head is elected by the following rules,

Rule 1: A host declares itself a cluster-head if its own host ID can be received during Floodmin rounds.

Rule 2: Otherwise, if a host can find that the same host IDs occurs at least once as WINNER[3] in both Floodmax and Floodmin, the one with minimal metric is selected as cluster-head.

Rule 3: Otherwise, the host which has the maximal metric in the k rounds of Floodmax, is elected as cluster-head.

[3] WINNER in KCMBC is the host which has the winning compound metric for a particular broadcast round

Since Floodmax is a greedy algorithm and may results in an unbalanced loading for cluster-heads, the minimal metric in the second rule for cluster-head election is used to promote fairness and distribute the load among the elected cluster-heads. Through the above procedure, host h_i becomes a cluster-head, if it has the largest the compound metric in k-hop neighborhood, or it has the largest the compound metric in the k-hop neighborhood of at least one host h_j.

In the above procedure, either Floodmax or Floodmax algorithm makes each host propagates the host metrics for k rounds. Consequently, each host initiates $2k$ control packets to elect cluster-heads. In an ad hoc network with N mobile hosts, the total overheads created by Floodmax and Floodmax algorithms are $2kN$ packets, which are much less than the number of overhead packets generated by using the CBKC algorithm [81]. A comparison of KCMBC and CBKC in terms of number of overhead packets is presented in Section 3.4

Moreover, the correctness of KCMBC is solely dependant on hosts electing cluster-heads that actually become cluster-heads. The proof of the correctness can be found in Appendix C.

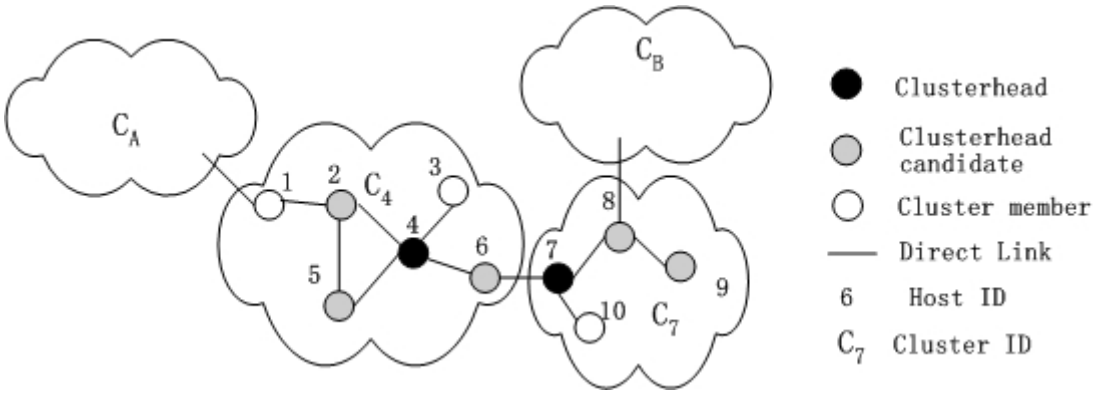


Figure 3-3. 2-hop cluster formation using KCMBC

Figure 3-3 shows an example of 2-hop clusters formed using KCMBC. The winning metrics of each broadcast round are listed in Table 3-1. From this table, it can be seen how the two hosts h_4 and h_7 are elected as cluster-heads. It is supposed that the average link available time of host 1, 3 and 10 is less than the predefined threshold T_{ALT}, then those hosts cannot be cluster-head candidates. Their winning metrics are initialized as $CP = (0,0,0)$. Moreover, as the average link expiration time $T_7 > T_8$, i.e. $CP_7 > CP_8$, host 8 cannot be selected as cluster-head, although it has the same number of neighboring hosts as host 7.

Table 3-1. Winning metrics for cluster-head selection $k = 2$

Host ID / Stage		1	2	3	4	5	6	7	8	9	10
Initialize		(0,0,0)	(3,5.6,2)	(0,0,0)	(4,4.8,4)	(2,6.1,5)	(2,5.4,6)	(3,5.1,7)	(3,4.2,8)	(1,3.9,9)	(0,0,0)
Floodmax Round	1	(3,5.6,2)	(4,4.8,4)	(4,4.8,4)	(4,4.8,4)	(4,4.8,4)	(4,4.8,4)	(3,5.1,7)	(3,5.1,7)	(3,4.2,8)	(3,5.1,7)
	2	(4,4.8,4)	(4,4.8,4)	(4,4.8,4)	(4,4.8,4)	(4,4.8,4)	(4,4.8,4)	(4,4.8,4)	(3,5.1,7)	(3,5.1,7)	(3,5.1,7)
Floodmin Round	1	(4,4.8,4)	(4,4.8,4)	(4,4.8,4)	(4,4.8,4)	(4,4.8,4)	(4,4.8,4)	(3,5.1,7)	(3,5.1,7)	(3,5.1,7)	(3,5.1,7)
	2	(4,4.8,4)	(4,4.8,4)	(4,4.8,4)	(4,4.8,4)	(4,4.8,4)	(4,4.8,4)	(3,5.1,7)	(3,5.1,7)	(3,5.1,7)	(3,5.1,7)
Cluster-head		4	4	4	4	4	4	7	7	7	7

3.3.3 Converge-Cast

In the Max-Min heuristic, after the cluster-head is determined, gateway initiated converge-cast is used to convey the cluster membership to this cluster-head. This converge-cast starts from gateway hosts. Each intermediate host adds its own information to the converge-cast packet and forwards this packet along the path to the elected cluster-head. However, the heuristic may lose the non-gateway hosts which are located on the cluster border but have no link to connect neighboring clusters. For example, the border hosts 3, 9 and 10 in Figure 3-3 cannot deliver their membership to cluster-heads via gateway initiated converge-cast.

To overcome the above shortcoming, the KCMBC approach employs a distance related converge-cast instead. During each round of Floodmax and Floodmin, the shortest distance (the number of hops) to the corresponding WINNER is attached in the broadcasted packet. Every host records this shortest distance and the SENDER host [4] from which the broadcasting packet containing the winning metric has been heard. After the host determines its cluster-head, it knows the shortest distance to the selected cluster-head, and broadcast this information as well as the ID of its cluster-head to all of its neighboring hosts. Consequently, each host knows which neighboring hosts have longer distance to the same cluster-head.

The distance related converge-cast begins from the border hosts which have the longest distance to the cluster-head in the neighborhood. Converge-cast packets deliver the host ID, the cluster-head's ID and the list of neighboring hosts (including the neighboring clusters connected to gateways). On the other hand, those non-border hosts wait to receive all the packets from its neighboring hosts which have longer distance to the same cluster-head. After all of such neighboring hosts have been heard, the host adds its cluster information to the converge-cast packet and rebroadcasts it. The process continues until the cluster-head has received packets from all of its neighboring hosts. For example, as shown in Figure 3-3, the transmission order of the distance related converge-cast within cluster C_4 should be: $h_1 \rightarrow h_2 \rightarrow h_4$, $h_5 \rightarrow h_4$, $h_3 \rightarrow h_4$ and $h_6 \rightarrow h_4$.

After the converge-cast is completed, the cluster-head holds the list of membership in its cluster, which is broadcasted to all cluster members along the inverse direction of the converge-cast. Consequently, the whole process of cluster formation in the KCMBC approach can be summarized by the flow chart as shown in Figure 3-4.

---

[4] Once the cluster-head has been selected, SENDER is used to determine the shortest path back to the cluster-head.

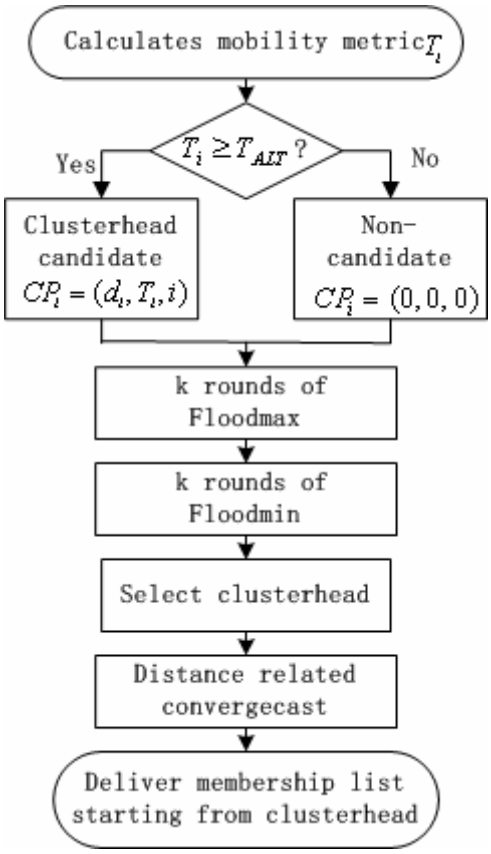


Figure 3-4. Flow chart of the KCMBC approach

3.3.4 Cluster Maintenance

In a dynamic ad hoc network, when a host disconnects from or connects to its neighboring hosts, the cluster structure may be altered. The frequent changes of cluster structure generate much control overhead, which may adversely affect the network performance and consume the scarce bandwidth. Therefore, it is important to design a cluster maintenance scheme to keep the cluster structure as stable as possible. Comparing to other cluster maintenance schemes [40][78][81]. The KCMBC approach applies a more effective mechanism to cope with host/link activation and host/link deactivation.

A. Host/link Activation

The primary objective of an activating host is to discover an adjacent host and join a cluster. Each mobile host in the network constantly monitors the links connecting with its neighboring hosts through periodical beacons. When an orphan host h_j either moves close to an existing cluster or switches on its power in a cluster, the event of host activation is triggered. With the aid of the new connected neighboring host, the orphan host h_j checks whether the distance between h_j and the nearest cluster-head is no more than k hops. If the constraint is satisfied, h_j tries to join this cluster. It is possible that h_j has chance to join more than one clusters simultaneously. For example, h_j has new links to connect both h_m and h_n, which belong to different clusters. In this case, according to equation (27), h_j calculates the link expiration time T_{im} and T_{in}. Then h_j joins the cluster which owns the host with the larger link expiration time. On the other hand, a group of orphan hosts may be unable to join existing clusters due to sparse host density. In this case, once more than d (the average host degree in the network) orphan hosts are detected in the neighborhood, those orphan hosts attempt to trigger a new cluster formation process.

Link activation is caused either by host activation or by newly created wireless link because of host mobility. There are several cases to consider when a new link is created between two hosts h_i and h_j. Firstly, if neither of h_i and h_j is a cluster-head, no change in the cluster structure is made. Secondly, if an ordinary cluster member h_j moves into the range of another cluster-head h_j, cluster re-affiliation is not triggered. Lastly, if both h_i and h_j are cluster-heads, reclustering is triggered. In this case, the cluster-head with the less average

link expiration time will dismiss its cluster, and hosts of this cluster will join the winning cluster-head.

B. Host/link deactivation

On the other hand, host deactivation occurs when a host leaves from its cluster, or its power is switched off. Under KCMBC, if a cluster member h_j loses the path to connect the original cluster-head, or the distance between the cluster-head and h_j is larger than k hops, h_j will depart from the cluster and try to join other clusters. Host deactivation or link outage may make the cluster structure worse. Clusters with poor structure, which only contain cluster-heads and gateways, will be dismissed.

Furthermore, cluster-head change occurs if a cluster-head h_i cannot work properly, when h_i wants to switch itself off, or h_i predicts that its last neighboring host will be disconnected quickly based on the predicted link expiration time. In this case, according to the current member list, h_i selects a new cluster-head which can connect the most cluster members. Those hosts which cannot be included in the new k -hop cluster will try to join other clusters nearby. Consequently, the control overhead caused by cluster reconstruction is minimized. An example of cluster-head change is shown in Figure 3-5, where the original cluster-head (host 5) will switch itself off. A new cluster-head (host 8) is selected and all cluster information is copied from host 5 to host 8. Moreover, since the distance between host 4 and the new cluster-head is larger than $k = 2$, host 4 becomes an orphan host, but the other previous cluster members can be included in the new cluster C_8

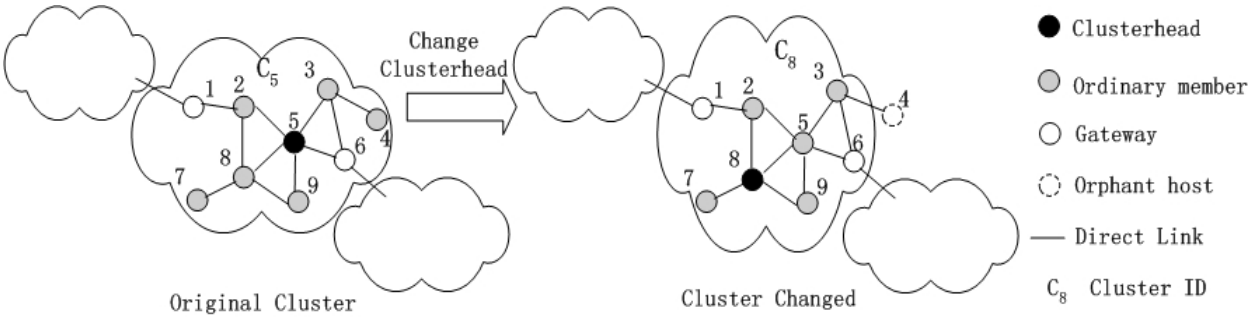


Figure 3-5. The change of Cluster-head, $k = 2$

3.4 Performance Evaluation

To evaluate the performance of KCMBC, simulation experiments are conducted in a homogeneous network environment. Every host has a unique host ID and a transmission range of radius $r = 1$ unit. Hosts are randomly distributed in an area of $S = 30 \times 30$ square units. The number of hosts (N) in the network ranges from 1000 to 4000. According to equation (1) in Chapter 2, the average host degree d varies from 3.5 to 14.0. The maximum number of hops (k) between a cluster member and the cluster-head is set from 1 to 4. Every host moves according to a uniformly distributed moving rate among the range of $[0, V_m]$. The maximum host moving rate V_m is set as 0.5 Unit/Sec unless it otherwise stated. Through beacons, each host broadcasts its current position to the neighboring hosts once every second. Each simulation experiment runs 1000 seconds of simulation time. Moreover, in the simulations for KCMBC, the predefined threshold for cluster-head candidates is set as $T_{ALT} = 0.8E(T_{X1})$ based on experimentation, where $E(T_{X1})$ is the average link available time in the network.

The performance of the proposed KCMBC approach are compared with other three k-hop clustering schemes, i.e. the Max-Min heuristic (Max-Min), the CBKC scheme [81] based on

lowest ID (CBKC_ID) and the CBKC scheme based on highest degree (CBKC_Degree). In the CBKC_ID, the host with the highest ID in the k-hop neighborhood is elected as the cluster-head. In the CBKC_Degree scheme, the host with the highest connectivity in the k-hop neighborhood is elected as the cluster-head. The performances of each clustering scheme are evaluated, in terms of the number of cluster-heads, the average cluster size, the variance of cluster size, cluster-head duration and the overheads for cluster formation.

Figure 3-6 shows the average cluster size versus the number of hosts (N) in the network. When k and N are fixed values, the two CBKC schemes produce the largest cluster sizes followed by the KCMBC approach and finally the Max-Min heuristic. It is clear that the cluster size for KCMBC is larger than that for Max-Min. This is because that KCMBC takes into account the connectivity of hosts during cluster formation. Likewise, the cluster size for CBKC_Degree is larger than that for CBKC_ID.

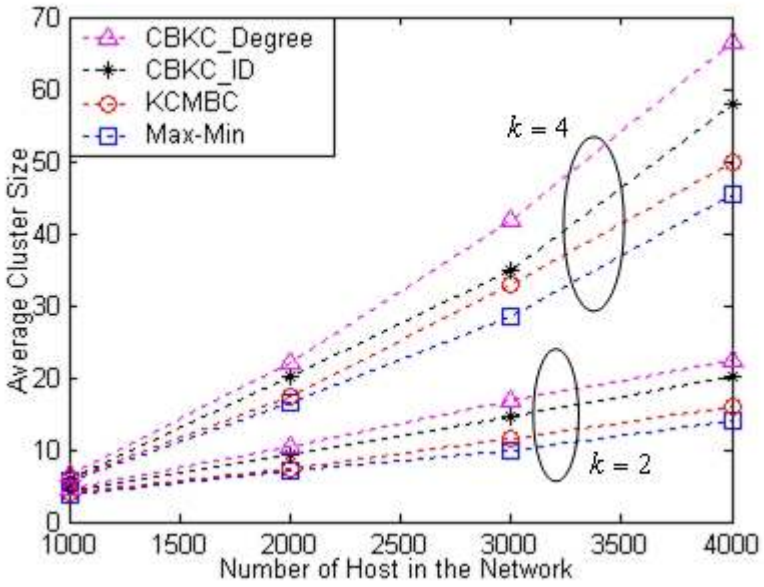


Figure 3-6. The average number of hosts in a cluster

From Figure 3-6, it can be found that for each clustering approach, the cluster size largely increases with k, and the cluster size for $k=4$ increase faster than that for $k=2$. For example, using KCMBC, when $N=4000$, the cluster size for $k=4$ is about 4 times of that for $k=2$. It is not expected that the average size of a cluster is too large, as clusters will be overloaded with too many cluster members. Nor is it good that the cluster size is too small, since the efficiency of mobility management based on cluster structure will largely decrease. To balance this trade off, the selection of k should consider both the bandwidth resource of each host and the diameter of network. A suitable value of k can be set in the range from 2 to 4.

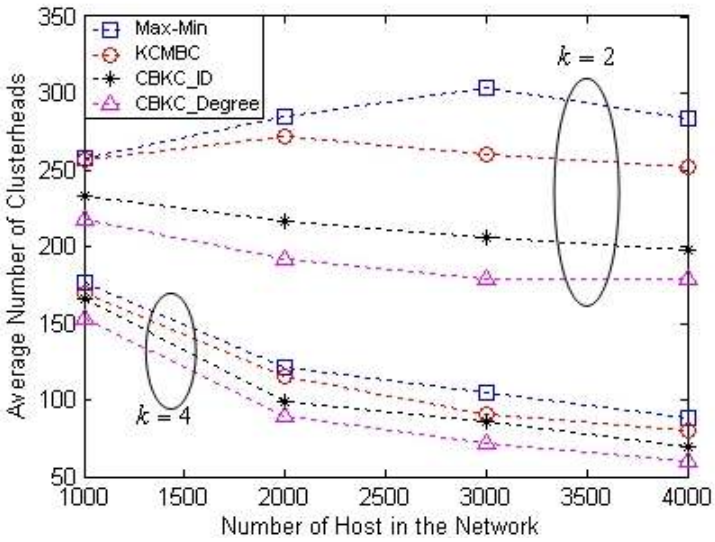


Figure 3-7. The average number of cluster-heads in the network

Figure 3-7 illustrates the average numbers of cluster-heads for the four schemes. It can be seen that the two CBKC schemes generate the least number of cluster-heads, and the average number of cluster-heads for KCMBC is less than that for Max-Min. It also can be found when N has a fixed value, the number of cluster-heads largely decreases with the increase of k, since the average cluster size increases with k, and the increase of cluster size leads to

the decrease of number of cluster-heads when N has a fixed value.. Accordingly, a large value of k can make the cluster structure scalable for large and dense networks.

Table 3-2. The variance of cluster size

(a). $k = 2$

Scheme \ N	1000	2000	3000	4000
KCMBC	11.78	35.64	78.99	138.78
Max-Min	12.34	37.29	82.56	152.78
CBKC_ID	16.61	81.96	206.14	399.83
CBKC_Degree	20.17	105.42	274.28	490.62

(b). $k = 4$

Scheme \ N	1000	2000	3000	4000
KCMBC	27.16	202.38	514.26	1184.13
Max-Min	29.16	217.38	548.92	1242.75
CBKC_ID	36.61	407.78	1221.17	3354.79
CBKC_Degree	44.20	495.44	1736.47	4428.45

On the other hand, the variance of cluster size indicates the size uniformity of clusters. A small variance of cluster size represents that most clusters in the network have uniform cluster size, and few orphan hosts are generated. Contrarily, a large variance of cluster size means that the sizes of clusters are irregular and difficult to be managed. Table 3-2 lists the variance of cluster size for the four clustering schemes with $k = 2$ and $k = 4$, respectively. It is clear that variance of cluster size for KCMBC and Max-Min is much smaller than the two CBKC schemes. This is because that KCMBC and Max-Min apply the similar procedure to construct load balancing clusters. Contrarily, the greedy algorithm deployed in the two CBKC schemes cannot generate load balancing clusters. Compared with Max-Min, KCMBC has smaller variance of cluster size. Under KCMBC, the distance related converge-cast guarantees that all border hosts can join their selected cluster-heads. By contrast, using Max-Min, some border hosts may lose the opportunities to join clusters. Therefore,

comparing to KCMBC, Max-Min generates more orphan hosts and causes the variance of cluster size larger.

To compare the performance of the two different convergence mechanisms applied in KCMBC and Max-Min approaches, the number of lost cluster members is defined as the number of hosts in the network which cannot join the selected clusters due to the convergence mechanism. Table 3-3 shows the number of lost cluster members for $k = 2$ and $k = 4$, respectively. It can be found that this number for KCMBC always keeps zero, because the distance related converge-cast applied in KCMBC is able to correctly deliver the information of every cluster member to its cluster-head. On the contrary, the number of lost cluster members for Max-Min is always more than zero. From Table 3-3, it can be seen that the number of lost cluster members for Max-Min increases with the decrease of N. This is because when host density becomes sparser, there are many border hosts lacking the links to connect neighboring clusters in the network, which cannot send their information to the elected cluster-heads by using the gateway initiated converge-cast of Max-Min.

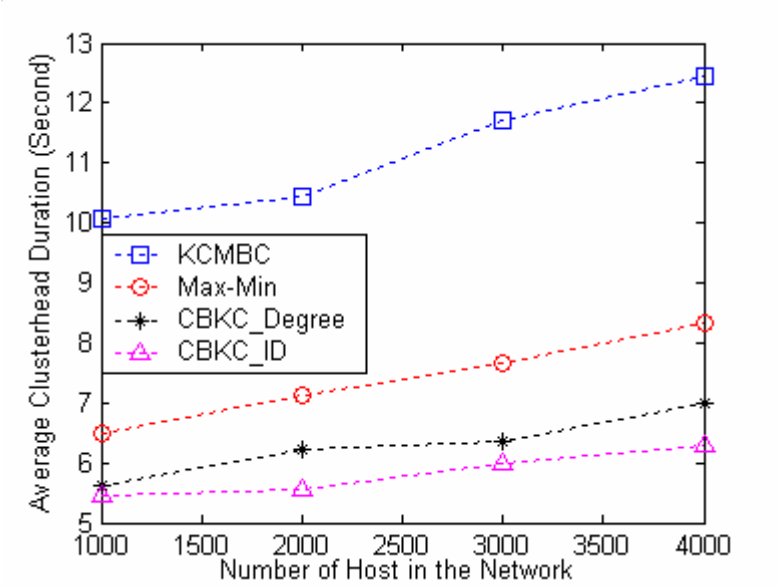
Table 3-3 The number of lost cluster members

Scheme \ N		1000	2000	3000	4000
$k = 2$	Max-Min	87	52	36	21
	KCMBC	0	0	0	0
$k = 4$	Max-Min	112	71	45	28
	KCMBC	0	0	0	0

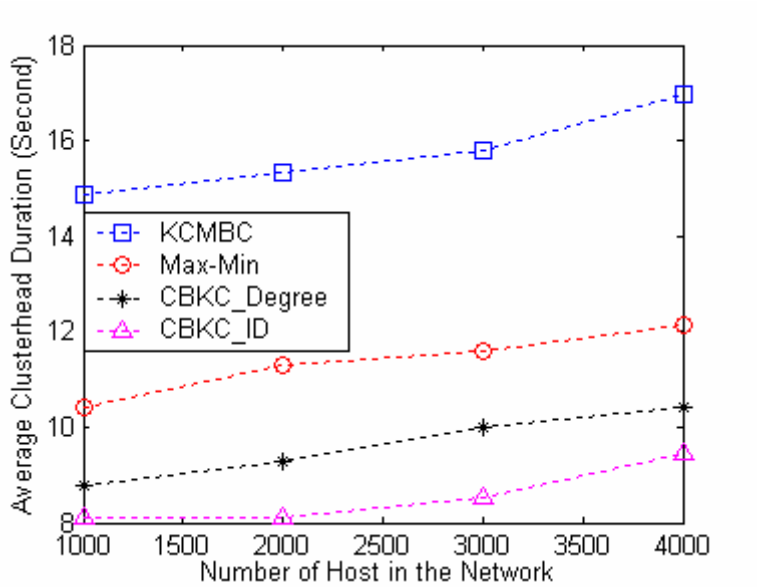
The stability of cluster structure using different schemes is evaluated by cluster-head duration, which is defined as the time of a host being a cluster-head once it is elected as a cluster-head. Long average cluster-head duration indicates that the cluster structure is stable and the frequency of cluster reconstruction is low. The average cluster-head durations for the four schemes are measured by simulations, where the cluster maintenance for KCMBC

applies the mechanism as described in Section 3.3.4, and the cluster maintenance for the other three schemes deployed the same mechanism used in literature [81].

Figure 3-8 (a) and (b) shows the average cluster-head duration for $k = 2$ and $k = 4$, respectively. It can be seen that KCMBC has the highest cluster-head duration, followed by Max-Min and finally the CBKC_Degree and CBKC_ID. Compared with Max-Min, KCMBC increases the average cluster-head duration up to 50%. This is because that in KCMBC, the average link expiration time is used to be mobility metric for cluster-head election. It is also noticed that the use of the cluster maintenance mechanism of KCMBC is able to increase the cluster stability. Figure 3-8 shows for the each scheme, the average cluster-head duration slightly increases while the number of mobile hosts in the network increases. From the comparison of Figure 3-8 (a) and (b), it can be found the cluster-head duration for $k = 4$ is higher than that for $k = 2$. This is because that the increase of k leads to the increase of the cluster size and the decrease of the frequency of reconstructing clusters.



(a) $k = 2$



(b) $k = 4$

Figure 3-8. The average cluster-head duration

$k = 2$, $V_m = 0.5$ Unit/s

Figure 3-9 illustrates the effect of the maximum host moving rate V_m on the average cluster-head duration, when $k = 2$ and $N = 2000$. For each clustering scheme, the cluster-head duration decreases with the increase of V_m, since the cluster topology changes rapidly when host moving rate becomes large. From Figure 3-9, it also can be seen that the cluster-head duration for KCMBC is always the highest among the four clustering schemes when V_m has a fixed value. Compared with Max-Min, the proposed KCMBC scheme significantly improves the cluster stability in dynamic environment. It can be concluded that KCMBC is highly suitable for creating stable clusters in the network with variable host mobility.

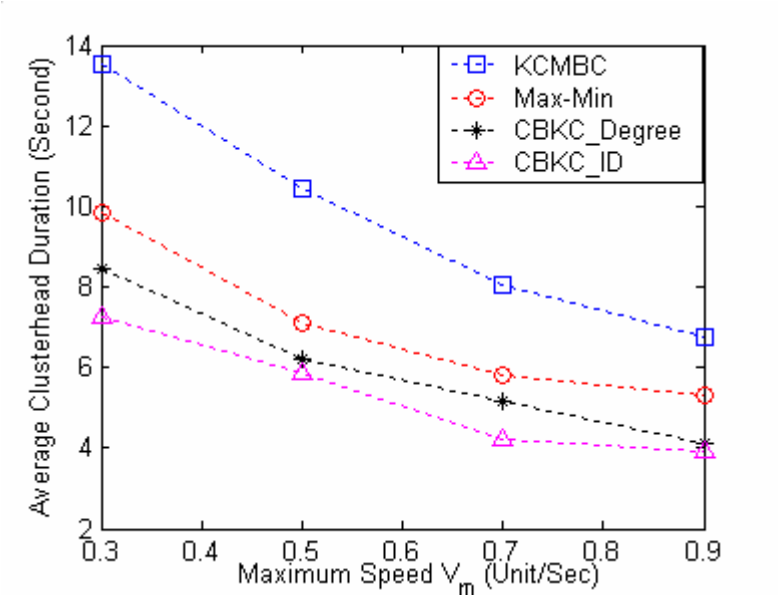


Figure 3-9. The impact of host mobility on cluster-head duration

$k = 2$, $N = 2000$

On the other hand, the scalability of a clustering approach depends highly on the control overheads created by cluster formation, which is denoted as O_{CF}. It is known that the values of O_{CF} in KCMBC and Max-Min are the same, and the values of O_{CF} in CBKC_ID and CBKC_Degree are also the same. Therefore, we only need to focus on the comparison of the values of O_{CF} between KCMBC and CBKC_ID.

In the KCMBC approach, each host initiates $2k$ rounds of Floodmax and Floodmin to elect cluster-heads. Once the cluster-head is determined, each host broadcasts one declaration packet to its neighboring hosts and delivers one converge-cast packet. In addition, every cluster member forwards the cluster membership list which is sent by the cluster-head. Hence, the total overheads for cluster formation in the network are given by

$$O_{CF\_KCMBC} = (2k+3)N \text{ packets.} \quad (29)$$

In the CBKC scheme based on lowest ID, each host must broadcast its host ID to all its k-hop neighboring hosts. The host, which has the lowest host ID in the k-hop neighborhood, broadcasts the decision to create cluster. After all the k-hop neighboring hosts with lower IDs have been heard from, each host broadcasts its clustering decision only once. Let N_k denote the average number of k-hop neighboring hosts of each host. The total overheads created by cluster formation can be obtained as

$$O_{CF\_CBKC} = 2N_k N \text{ packets.} \tag{30}$$

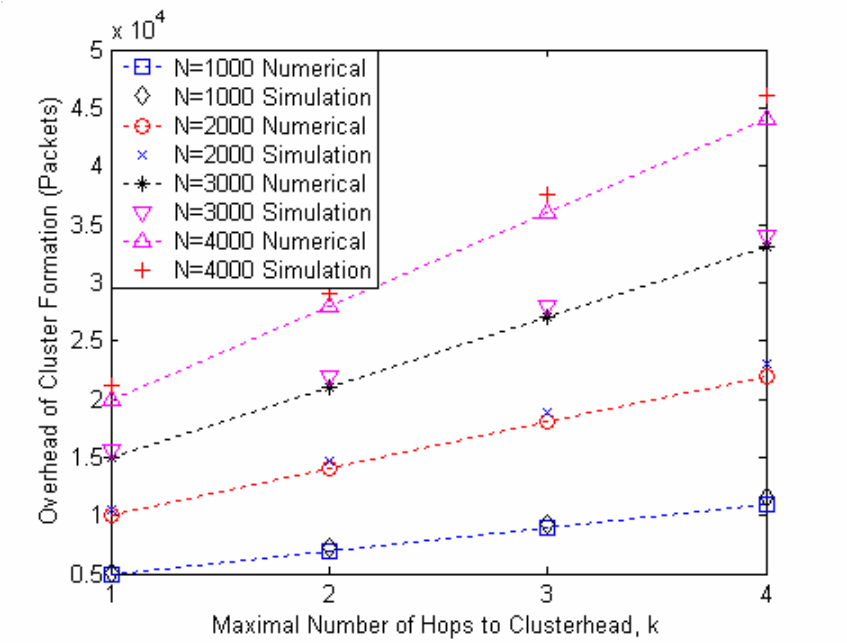


Figure 3-10. The overheads for cluster formation using KCMBC

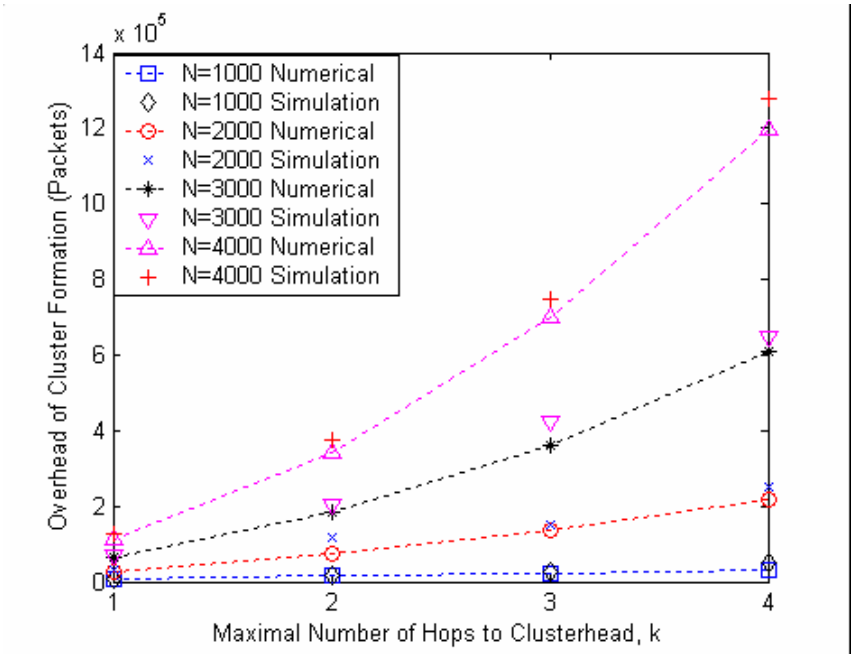


Figure 3-11. The overheads for cluster formation using CBKC

The above analysis on overheads is validated by simulations. Figure 3-10 shows the simulation results and the numerical results of the overheads created by cluster formation using KCMBC, and Figure 3-11 show the overheads created by cluster formation using CBKC. The numerical results in the two figures are predicted by equation (29) and equation (30), respectively. The figures illustrate that the simulation results match the predicted results well. From both Figure 3-10 and Figure 3-11, it can be seen that the increase of either k or N results in the increase of control overheads. Moreover, when k and N have fixed values, O_{CF_KCMBC} is much less than O_{CF_CBKC}. For a fixed value of N, O_{CF_KCMBC} increase with k following a linear rate, however, O_{CF_CBKC} increases with k following approximately a square rate. When $k=4$ and $N=4000$, O_{CF_CBKC} is about 26 times of O_{CF_KCMBC}. Consequently, the control overheads generated by KCMBC is much smaller than those generated by CBKC.

According to the evaluation results in this section, it can be concluded that among the four k-hop clustering approaches. The proposed KCMBC approach is able to generate clusters with modest size and the smallest variance of cluster size, and the average cluster-head duration for KCMBC is the longest. The cluster structure created by KCMBC is more stable than that created by other approaches. While the two CBKC schemes have larger cluster size than KCMBC, they suffer greatly in other statistics such as variance of cluster size and cluster-head duration. Moreover, the overheads created by cluster formation using the KCMBC scheme are much less than those using the CBKC schemes.

3.5 Conclusion

In order to keep cluster structure robust in a dynamic ad hoc network with large scale, a novel k-hop Compound Metric Based Clustering (KCMBC) approach is proposed in this chapter. Since KCMBC inherits the cluster formation procedure of the Max-Min heuristic, it has small control overhead, fast convergence speed and good scalability for large scale network. On the other hand, KCMBC applies some innovations to improve the performance of clustering: (1) the average link expiration time of each host is used to characterize the degree of mobility of a host with respect to its neighboring hosts; (2) only the cluster-head candidates, which satisfy the condition of host mobility, can be elected as cluster-heads; (3) the compound metric combining host connectivity, the average link expiration time and host ID is used to select cluster-heads; (4) the distance related converge-cast takes the place of gateway initiated converge-cast to gather the information of cluster members. (5) an efficient cluster maintenance mechanism is employed to cope with local topology changes. As KCMBC does not restrict the number of gateways, more than one inter-cluster links may connect two neighboring clusters. This redundant cluster architecture can tolerate false route

and balances traffic load. Simulation experiments show that clusters created by the KCMBC approach have modest but more uniform cluster size. Comparing to other clustering schemes, KCMBC has the longer cluster-head duration, which indicates that the cluster stability can be largely increased by KCMBC. With the above significant advantages, the KCMBC approach is able to provide a robust infrastructure, which is not easily disrupted by host mobility and can support hierarchical routing, location service and mobility management.

Chapter 4 The Location Service Protocol Based on K-hop Clustering

4.1 Introduction

Location information of mobile hosts has been used to improve the performance of routing protocols in mobile communication networks [28][29][30][31][32]. A mobile host is able to obtain its location information either by GPS receivers or by some other type of position service [33][34]. In order to make the location based routing protocol operating properly, each mobile host in the network must know the location of the neighboring hosts to relay packets. However, in most conventional mobile networks, the location based routing protocols are operating under the condition that the location information of all mobile hosts is always known already. In this case, location service protocol is required in order to find the physical location of individual host as well as the logical affiliation that mobile hosts register their current location with service.

In classic cellular networks, some special hosts are defined as location servers such as Home Location Register (HLR) to maintain the location information for the other mobile hosts, which register the current location to the dedicated location servers. When a host wishes to know the position of a desired communication partner, it contacts one of the location servers to obtain the required location information. However, the traditional HLR-like schemes are not applicable here since every mobile host in an ad hoc network is the peer entity with the same equipment and the random walking of hosts affects the accuracy of location information. In this case, the location server may have to be switched from host to host and

the location databases in server may have to be updated frequently to match the random walking of mobile hosts.

The efficiency of location service protocols depends highly upon the availability of timely and accurate location information. Since frequent topological changes are expected in ad hoc networks, the distribution of up-to-date information can easily saturate the network. On the other hand, location updates arriving late due to latency can drive network routing into instability. As one part of routing protocols, a few location service schemes have been proposed in recent years. Technical reviews on existing location service approaches are presented in [60] and [71].

The Distance Routing Effect Algorithm for Mobility (DREAM) [28] is used to provide the location information that every mobile host in the network maintains and updates regularly the position information for the whole network using flooding method. In this case, the accuracy of the position information depends on (1) the frequency of updating position information, which depends on the mobility degree of a host and (2) convergence time for flooding the position update packet across the network, in which the distance effect [54] has be taken into account that the direct neighboring hosts from the flooding source have more accurate position information than that of hosts with long distance from the flooding source. On the other hand, the communication complexity of the location information update and maintenance is in scales of $O(N)$, where N is the number of hosts in the network. In general, the DREAM scheme is inappropriate for large scale ad hoc networks.

The Zone-based Hierarchical Link State (ZHLS) routing protocol [82] is another protocol to provide the location information in mobile networks, where hosts determine the zone ID by

mapping the physical location to a non-overlapping zone map so that hosts know the connectivity within the same zone and also the zone level network topology. Although ZHLS generates fewer overheads than the schemes based on flooding, the implementation is not flexible for dynamic networks such as ad hoc networks, since the zone map has to be assigned in the design stage. On the other hand, the performance of ZHLS may be depressed by the long latency and the numerous overheads for location search.

The Grid Location Service (GLS) [32] is an efficient scheme, which relies on a grid-based hierarchy overlaying the network area. Under this scheme, a large square area is recursively divided into relatively smaller square sub-areas. In this case, the number of location servers increases logarithmically as the number of hosts increases but the corresponding communication complexity still remains at $O(\sqrt{N})$. Unfortunately, the application of GLS is usually limited since the network has to be partitioned into a hierarchy of grids. On the other hand, when network mobility is group based, such as military troops moving in battlefield [83], the network cannot be simply divided into predefined grids. Improper partition may result in resource wastage and network performance decrease. A similar location service scheme for doubling circles routing protocol [77] applies circles, instead of square grids.

The Quorum based location service is evolved from information replication in databases and distributed systems. Haas and Liang [84] propose a quorum based distributed mobility management scheme based on virtual backbone, which are database servers for location information. The generation of a virtual backbone was considered as a Minimum Set Covering (MSC) problem, and a Distributed Database Coverage Heuristic (DDCH) are presented in [37]. One of the advantages of this scheme is that the heuristic is able to

dynamically maintain the structure of virtual backbone as the network topology changes. However, the drawback of the virtual backbone system is that the mobility management generates a large amount of data overheads for network control, since every host must update and maintain the link states in its r-hop neighborhood. However, if the destination cannot be found in the local area, then the sender host sends a location paging query across the backbone and waits for query response from the destination host. Of course, latency will be introduced in such location searching process.

Home agent scheme [85][86] is a special case of quorum system, in which mobile host selects a home region and then sends the location updates to the other hosts currently located in the home region. In order to locate destination, hosts send a destination search packet toward the home region of destination. The position of the home region for a host can be derived by applying the hash function based on host ID. However, the maintenance of home agent for each mobile host is not considered in the proposed schemes. In this case, mobile hosts must be remained in a relatively fixed region. Moreover, the total overhead of managing home agent is unknown in a dynamic environment.

However, there is no any location service scheme with low network control overheads that has been adaptive in dynamic mobile networks so far. The main objective of this chapter is to design an efficient location service protocol, which has not only good scalability but also the ability of self-organization and strong adaptability to accommodate most applications of ad hoc networks. Inspired by the distance effect [54], the virtual backbone quorum scheme and the ZHLS protocol, the proposed k-hop Clustering Based Location Service (KCBL) protocol is able to provide location service and mobility management in a large scale ad hoc network.

4.2 The KCBL Protocol

It is well known that a hierarchy is an effective architecture to achieve adequate scalability in very large networks. In fixed infrastructure networks, hierarchical aggregation achieves the effect of making a large network appear much smaller from the perspective of routing algorithm [87][88]. When there is no fixed infrastructure available such as in ad hoc networks, the most popular way of building hierarchy is to group hosts geographically close to each other into clusters. Cluster-based routing in an ad hoc network can also make a large network appear smaller, but it is more important that Cluster-based routing protocol is able to make the network topology to be less dynamic. Differing from other strategies of grouping mobile hosts using fixed geographic scopes such as grid [32] or circle [77], the cluster-based routing protocol in an ad hoc network is self-organizing and adaptable to build a mobility management system. Such cluster-based architecture can also perform many other network functions, which are not elaborated in this chapter, such as channel assignment, flow control , transmission power control and battery saving.

Location management based on cluster-based architecture can be applied to either single level clusters or multiple level clusters. In a multilevel clustering scheme such as the Hierarchical State Routing (HSR) [41][89], hosts located in the same logical level are grouped into clusters. The elected cluster-heads at the lower level become members of the next higher level. Similar to the Internet hierarchy, HSR has good scalability. However, the implementation of multiple-level cluster-based topology for location service in dynamic network is complicated. Firstly, due to host mobility, the reconstruction of hierarchical structure generates a large amount of overheads. Secondly, each host must maintain a

hierarchical address, which may be changed randomly when the clustering topology changes at any level. Such almost continuously changing of the hierarchical address makes it difficult to locate and to keep track of hosts. Thirdly, link activation/failure at lower level certainly affects the cluster structure at higher levels. The changing in hierarchical cluster membership is able to cause the reconstruction of the hierarchical location management tree. Lastly, since all packets are sent from source host to destination host through a set of backbone hosts, which comprise a centralized subnet, the high demand of channel bandwidth and host stability are imposed on these backbone hosts. Therefore, multilevel clustering is difficult to provide efficient location service with balanced loading on mobile hosts.

By contrast, the single level clustering is relatively easy to manage. Cluster-heads are dynamically elected among the mobile hosts to be the local coordinators. Each cluster-head maintains the membership of its cluster and tracks local topology changes due to host mobility. From the viewpoint of location management, cluster-heads can instinctively act as location servers which are distributed among all the mobile hosts in the network. Moreover, since the cluster size (the number of cluster members in the cluster) of k -hop cluster is more than that of 1-hop cluster, k - hop cluster structure has more scalability. A k -hop cluster C_m is a set of hosts with the same cluster-head h_m, and any cluster member is at most k hops far away from its cluster-head. As shown in Figure 4-1, every cluster consists of one cluster-head, ordinary cluster members which are located inside of a cluster, and gateways which are located on the cluster border and connect with neighboring clusters.

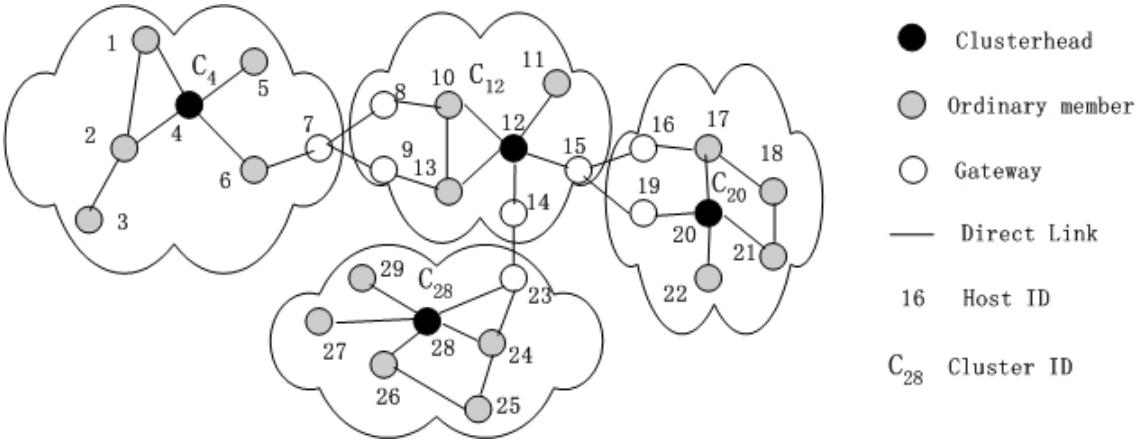


Figure 4-1. An example of k-hop clusters, k=2

The KCBL protocol is based on a single level cluster structure to provide location service. An ad hoc network can be divided into many non-overlapping clusters by either using a k-hop clustering scheme, such as Max-Min heuristic [55], KCMBC scheme (see Chapter 3) and CBKC approach [81], or by partitioning the network into several logical groups. When the network is partitioned into several logical groups, each group corresponds to a particular user team with common characteristics and the same mobility pattern, such as tank battalion in the battle field, search team in a rescue operation, moving behavior of the same company, or students within the same class. A logical group can consist of one or several k-hop clusters according to the number of group members.

4.2.1 The Overview of KCBL

It is supposed that each host has a unique host ID. The cluster ID is determined by the host ID of cluster-head. Under the KCBL protocol, the cluster-heads and the cluster members form a logical two level hierarchy. Cluster-heads work as distributed location servers, which

keep track of mobile hosts within the cluster and record the inter-cluster connectivity of the whole network. The cluster ID or the position of cluster-head is used to represent the location information of all the mobile hosts in this cluster. On the other hand, ordinary cluster members only hold the local topology information about theirs own cluster. When a host wants to obtain the location information of a destination which does not exist within the same cluster, it simply sends a location enquiry packet to its cluster-head. The cluster-head in turn immediately replies with the location response which is represented by the cluster ID of the destination or the location of the destination's cluster-head.

However, the cluster affiliation may be changed frequently due to dynamic movement of mobile hosts. In this case, the performance of location service depends highly on the efficiency of location management mechanism which should not only provide useful location information, but also keep the control overheads at a low level. Corresponding to clustering structure, location management in the KCBL protocol is performed on two levels: intra-cluster level and inter-cluster level.

Table 4-1. The LC table of the hosts in cluster C_{12} (shown in Figure 4-1)

Host	8	9	10	11	12	13	14	15
ID of Neighboring hosts	10, C_4	13, C_4	8,12,13	12	10,11,13,14,15	9,10,12	12, C_{28}	12, C_{20}
Distance to CH	2	2	1	1	--	1	1	1

Intra-cluster location management applies a link state routing approach [90] within a cluster. Once clusters have been constructed, each cluster member maintains a Local Connectivity (LC) table and an Intra-cluster Routing (IntraR) table. LC table [5] indicates the neighboring hosts of each cluster member, the neighboring clusters connecting with gateways and the

[5] At the initial stage, LC table is provided by the cluster-head, which knows the cluster topology through the cluster formation process.

distance to the cluster-head. For example, the LC table stored in the members of C_{12} as shown in Figure 4-1 is given by Table 4-1. When an intra-cluster link is broken or established, a Local Link Change (LLC) packet will inform all hosts within this cluster the new piece of link state information, and LC table stored by each cluster member will be updated accordingly. Moreover, based on the current LC table, every cluster member constructs an IntraR table by using a shortest path algorithm (such as Dijkstra's algorithm). As shown in Table 4-2, IntraR table indicates the next hop to reach a cluster member or a neighboring cluster. With the aid of IntraR table, an intra-cluster path or a route to one neighboring cluster can be easily established. Since link state exchange and routing table computation is localized, the intra-cluster location management is efficient and robust.

Table 4-2. The IntraR table of host 10 (shown in Figure 4-1)

Destination	8	9	11	12	13	14	15	C_4	C_{20}	C_{28}
ID of the next hop	8	13	12	12	13	12	12	8	12	12

Table 4-3. The LS table of host 4 (shown in Figure 4-1)

Cluster ID	C_4	C_{12}	C_{20}	C_{28}
Timestamp	T_{C_4}	$T_{C_{12}}$	$T_{C_{20}}$	$T_{C_{28}}$
Cluster-head Position	Coordinates of h_4	Coordinates of h_{12}	Coordinates of h_{20}	Coordinates of h_{28}
Member ID	1,2,3,4,5,6,7	8,9,10,11,12,13,14,15	16,17,18,19,20,21,22	23,24,25,26,27,28,29
Neighboring Clusters	C_{12}	C_4 , C_{20} , C_{28}	C_{12}	C_{12}

On the other hand, inter-cluster location management is operated by cluster-heads. Besides LC table and IntraR table, each cluster-head maintains one Location Service (LS) table, which describes the membership and inter-cluster connectivity of each cluster. In a LS table, each Cluster State (CS) item contains the cluster ID, the timestamp that indicates the time for generating the location information of corresponding cluster, the cluster-head's planar

coordinates (if available), cluster member list and the neighboring cluster list. Table 4-3 shows the LS table stored by host h_4 in Figure 4-1, which is the cluster-head of C_4.

4.2.2 Inter-Cluster Location Update

A cluster-head responds to location queries based on its LS table. In order to keep the accuracy of LS table in dynamic network environment, the location update rate must be fast enough to reflect topology changes. However, if all location update packets are propagated by using the ordinary methods such as flooding, a large amount of control overheads would consume scarce network bandwidth. Thus, an effective location update mechanism is required to increase the efficiency of inter-cluster location management and reduce control overheads.

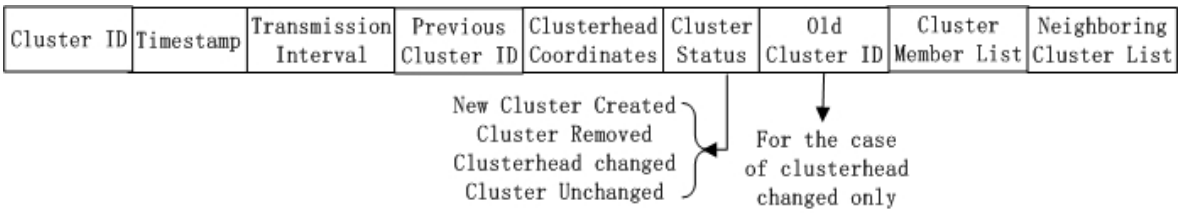


Figure 4-2. The layout of CS packet

Under the KCBL protocol, once a new cluster has been constructed, the cluster-head broadcasts a Cluster State (CS) packet, which carries its cluster information, to all other cluster-heads in the network. Based on the CS packets received, a cluster-head establishes its LS table. Moreover, CS packets are used to update inter-cluster location throughout the network. As shown in Figure 4-2, each CS packet represents the information of one cluster, including cluster ID, the timestamp (the CS packet generation time), transmission interval deferred by every cluster-head, the cluster-head's coordinates, the ID of the preceding cluster

from which the CS packet is heard, cluster status, cluster member list and neighboring cluster list. If a CS packet is generated by an existing cluster (i.e., the cluster status is "Cluster Unchanged"), the attached cluster member list and neighboring cluster list only contains the IDs of those cluster members and neighboring clusters which newly attach to or detach from this cluster. In this way, the redundant overheads are significantly reduced. Moreover, as the network spans a large area, a timestamp is used to eliminate duplicate copies and avoid delivery loops.

In a CS packet p_i, the attached transmission interval, denoted as τ_{pi}, indicates the time interval to be postponed by each cluster-head. τ_{pi} is used to control the propagation speed of this CS packet. For a new created cluster, τ_{pi} is set as zero, so the CS packet can be propagated immediately throughout the whole network, and the cluster initialization is able to be converged quickly. However, if the CS packet p_i is created by an existing cluster, τ_{pi} is larger than zero and packet p_i can only be propagated one cluster range every time interval τ_{pi}. This location updating mechanism is based on the distance effect [54]: since faraway mobile hosts appear to move more slowly than hosts nearby, it is not necessary for location server to maintain up-to-date location information on faraway hosts. Consequently, control overheads can be significantly reduced.

The transmission interval τ_{pi} can be determined by a predefined value. However, it is more reasonable that a cluster with highly dynamic structure should update its information frequently, and a cluster with stable membership need not update location information during a long time period. Hence, each cluster-head can set its own transmission interval based on the mobility pattern of this cluster or the group characteristics. Alternatively, considering that

the average link available time of each cluster is able to reflect the degree of the cluster membership changes, cluster-head h_i can simply set its transmission interval τ_{pi} by using:

$$\tau_{pi} = \sigma T_{HAVi} \tag{31}$$

where σ is the scaling factor, and T_{HAVi} is the average link available time of cluster C_i. According to the changes of LC table, the cluster-head can obtain T_{HAVi} by calculating the average link available time during a fixed time period T_{HM} ,.

As the local coordinator, every cluster-head maintains and monitors its cluster topology by using one cluster maintenance scheme [40][78][81]. A cluster-head h_i will generate and transmits its CS packet, if and only if one of the following events is detected during the past transmission interval τ_{pi}.

- Any member leaves from or joins to the cluster;
- A new neighboring cluster is connected or an old neighboring cluster is disconnected;
- The change of cluster topology, such as new cluster creation, cluster-head change or cluster removal.
- From the beginning that the previous CS packet is generated by the cluster-head h_i, the accumulated moving distance of h_i exceeds the predefined threshold D_{th}

On the other hand, upon receiving a new CS packet p_j which was created by cluster-head h_j, cluster-head h_i modifies the CS item for cluster C_j in its LS table according to the following rules.

- If the CS packet is created by a new cluster, a new CS item is created in the LS table;

- If the CS packet represents that an existing cluster has been dismissed, the corresponding CS item will be deleted from LS table;
- If the CS packet shows that one cluster C_K are newly added to or deleted from cluster C_j 's neighboring cluster list, i.e., the inter-cluster link between C_K and C_j is created or broken, cluster-head h_i will update the neighboring clusters of both C_j and C_K in LS table correspondingly.

The CS packet p_j will be deferred by cluster-head h_i for a transmission interval τ_{pj}, before it is forwarded to the neighboring clusters. In addition, according to the attached cluster ID and timestamp, each host checks if the CS packet has been forwarded. A cluster-head will discard the CS packet which has been forwarded, and a cluster member will deny the duplicate transmission of the identical CS packet destined for its cluster-head. Therefore, a CS packet is only relayed by each cluster once. The previous forwarding procedure continues until all cluster-heads in the network have received the packet p_j.

The propagation of CS packets can be implemented through an efficient multicast manner. Considering the fact that one gateway may connect more than one neighboring cluster simultaneously, each cluster-head can select its gateways to multicast CS packets destined for the neighboring clusters. The selection of forwarding gateways is the well-known Set Cover problem [91], which is NP complete. According to the LC table stored by the cluster-head, a greedy algorithm can be applied to select gateways as follows: the gateways connecting the more neighboring clusters but excluding the neighboring cluster from which the CS packet is received, have the higher priority to be selected. Such process is continued until all neighboring clusters are covered. The cluster-head only multicasts packets to those selected gateways through which the packets can be

delivered to the cluster-heads of neighboring clusters. Therefore, redundant transmissions can be largely reduced. For example, as shown in Figure 4-3, after receiving a CS packet from C_h, cluster-head h_i selects gateway h_1 and h_7 to multicast the CS packet to the neighboring clusters C_d, C_e, C_f and C_g.

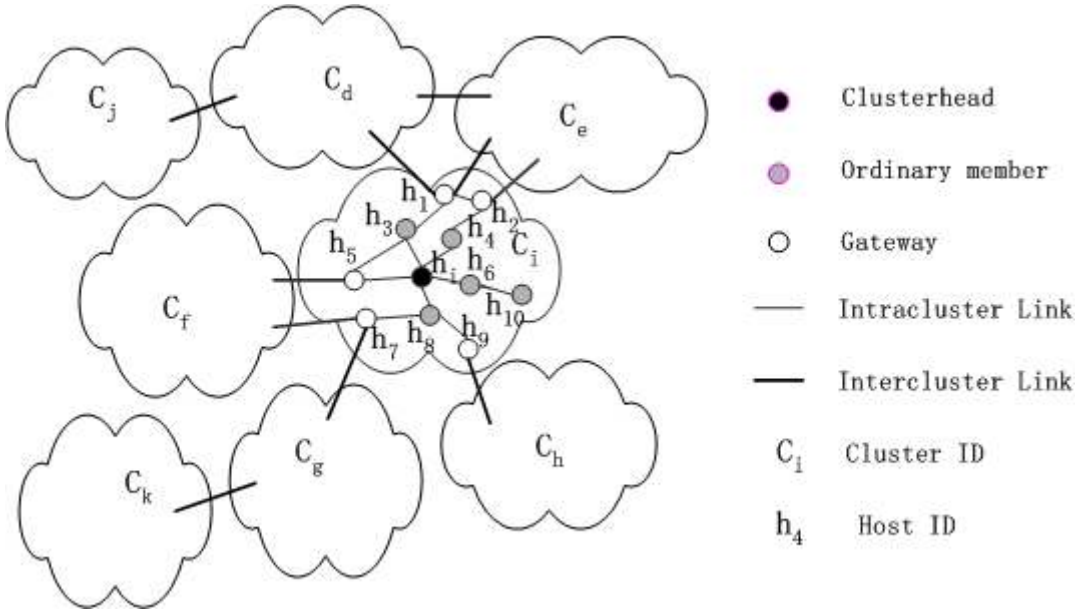


Figure 4-3. Gateway selection for multicast

4.2.3 Location Enquiry and Revision

Under the KCBL protocol, cluster-heads work as distributed location servers which provide location storage and retrieval. The location of a mobile host can be represented by its cluster ID or the position of its cluster-head, because of the following considerations. First, the cluster structure is more stable than a single wireless link. The host sojourn time in a cluster is relatively long, especially in a large cluster. Thus, the location information represented by a cluster can be valid for a long time period. Moreover, the host leaving from its previous cluster usually can be found within the neighboring clusters. Second, with the aid of IntraR table, an intra-cluster route can be quickly established and packets can be easily delivered

cluster by cluster. Therefore, the location information represented by a cluster can give vital information for routing search.

When the source h_i wants to initiate a communication path to the destination h_j, h_i sends a location enquiry packet to its cluster-head. The cluster-head in turn searches its LS table, and replies with h_j's location information, including h_j's cluster ID and the coordinates of h_j's cluster-head. If the source h_i is an orphan host which cannot join a cluster temporarily, the location enquiry will be forwarded to the nearest reachable cluster-head. In this case, the maximum latency of a response is only a round-trip transmission with $2(k+1)$ hops. Compared with other location service protocols [84][85][86] which must search the destination among several databases, the KCBL protocol has much shorter latency to respond location enquiries.

On the other hand, the KCBL protocol provides more accurate location information in the destination's neighborhood and less accurate information for the hosts far away. Let the cluster-hop distance denote the minimum number of clusters which should be traversed from the source to the destination. It is supposed that the transmission interval for a CS packet is τ_{pi}, then the latency of CS packet delivered to n cluster-hop distance away should be $T_{Cn} = n\tau_{pi}$. Therefore, if a long distance exists between the enquired cluster-head and the destination, the corresponding location information in the LS table may not be updated timely, and the destination may leave the cluster which is indicated by the LS table. However, the data packet can be sent properly even if the enquiry response provides slightly outdated location information. This is because from the view point of distance effect [54], a modest position variation appears to be very little for a host far away. Moreover, location

information can be revised by intermediate cluster-heads along the path to the destination, since the closer the packet reaches to the destination, the fresher location information can be obtained.

During a data packet delivery process, every packet carries the destination h_j's host ID and location information including the expected cluster ID, the coordinates of h_j's cluster-head and the timestamp. When the packet meets an inter-cluster link failure, or when the packet reaches h_j's expected cluster but h_j cannot be found within this cluster, this packet will be forwarded to the nearest reachable cluster-head. The cluster-head will update h_j's information labeled in this packet, if the timestamp for the corresponding CS item in the LS table is created after the attached timestamp. Then the route to the destination will be changed according to the new location information. This process is repeated until the packet reaches its destination.

4.2.4 Benefits to Routing Protocol

In the KCBL protocol, cluster-heads are dynamically selected among mobile hosts, and act as distributed location servers. Routing is still carried out in the flat network structure involving every host in the network, which allows balanced loading on the hosts and the links. Moreover, since each cluster-head only provide the location service to its cluster members and the orphan hosts in the neighborhood, location searching is more localized. The distribution of responsibility among cluster-heads provides a high degree of reliability in location service and mobility management. Even if a few cluster-heads become unavailable,

the failure of location service can be limited within the relevant clusters, and location updating and location querying are still possible in the remaining network.

On the other hand, the KCBL protocol can be used to increase the performance of routing protocols, in terms of reducing overheads, improving route stability and enhancing route recovery ability. In many proactive routing protocols such as Dynamic Source Routing (DSR) [44], each packet must carry a full path from the source to the destination. A large amount of overheads will be generated for a long routing path which spans a large scale network. By contrast, based on KCBL, a cluster level path is enough to provide a thorough route for any source-destination pair. The only modification for a routing protocol is that host level route is replaced by cluster level route, which can be obtained by sending the cluster-head a location enquiry. According to LS table, the cluster-head can determine a cluster level path to the destination, by using Shortest Path algorithm. The enquiry response carries the cluster level path as the form of a cluster list. Consequently, each data packet only needs to carry a cluster level path, which is much shorter than a host level path. With the aid of IntraR table stored by each host, the packet can be easily delivered from one cluster to the other, until it reaches to the destination.

As cluster architecture makes a highly dynamic topology appear much less changing, cluster level routes are more stable and adaptable for dynamic topology than common host level routes. Once a cluster level route has been established between a source-destination pair, the failure of one intra-cluster link scarcely influences the packet delivery, because the alternative path to reach the next cluster can be found immediately. On the other hand, if an inter-cluster connection is broken, the delivered packet will be forwarded to the nearest cluster-head for location information revising and route redesigning. Based on the LS table,

the cluster-head can design a new cluster level route to the destination. Therefore, with the service provided by the KCBL protocol, most link failure can be recovered rapidly. By contrast, in most existing routing protocols, a subsequent route searching has to be performed whenever the current route is broken. In this case, numerous overheads will be generated by a route recover process, and long transmission latency will be introduced.

Figure 4-4 illustrates an example for cluster level routing and location information revising. When host h_a in cluster C_i intends to communicate with host h_b, h_a sends a location enquiry to its cluster-head h_i. According the current LS table, h_i responds to the enquiry with h_b's cluster ID (C_h) and cluster-head h_h's coordinates. Moreover, this response provides a cluster level route between h_a and h_b, i.e., $C_i \rightarrow C_l \rightarrow C_m \rightarrow C_j \rightarrow C_g \rightarrow C_h$. Containing the cluster level path and h_b's location information, the forwarding packets are relayed cluster by cluster. It is supposed that the inter-cluster link between C_m and C_j is broken because of host mobility. Upon receiving a forwarding packet, the gateway of C_m finds that this packet cannot be delivered to C_j directly. In this case, the packet will be delivered to cluster-head h_m instead. h_m's LS table shows that h_b is still located inside of C_h, then h_m design a new partial route $C_m \rightarrow C_k \rightarrow C_j$ to replace the old one $C_m \rightarrow C_j$, and the packet delivery continues. When the packet is received by C_h's gateway, which detects the destination h_b does not exist in C_h at that moment. Then, the packet is forwarded to cluster-head h_h. Since h_h records that h_b has moved to C_e, the packet is delivered to C_e, and then it is forwarded to the destination h_b. At last, h_b sends the source h_a an acknowledgment packet, which contains the new cluster level route, i.e.,

$C_i \rightarrow C_l \rightarrow C_m \rightarrow C_k \rightarrow C_j \rightarrow C_g \rightarrow C_h \rightarrow C_e$. Subsequent packets from h_a to h_b will be delivered along the new cluster level route.

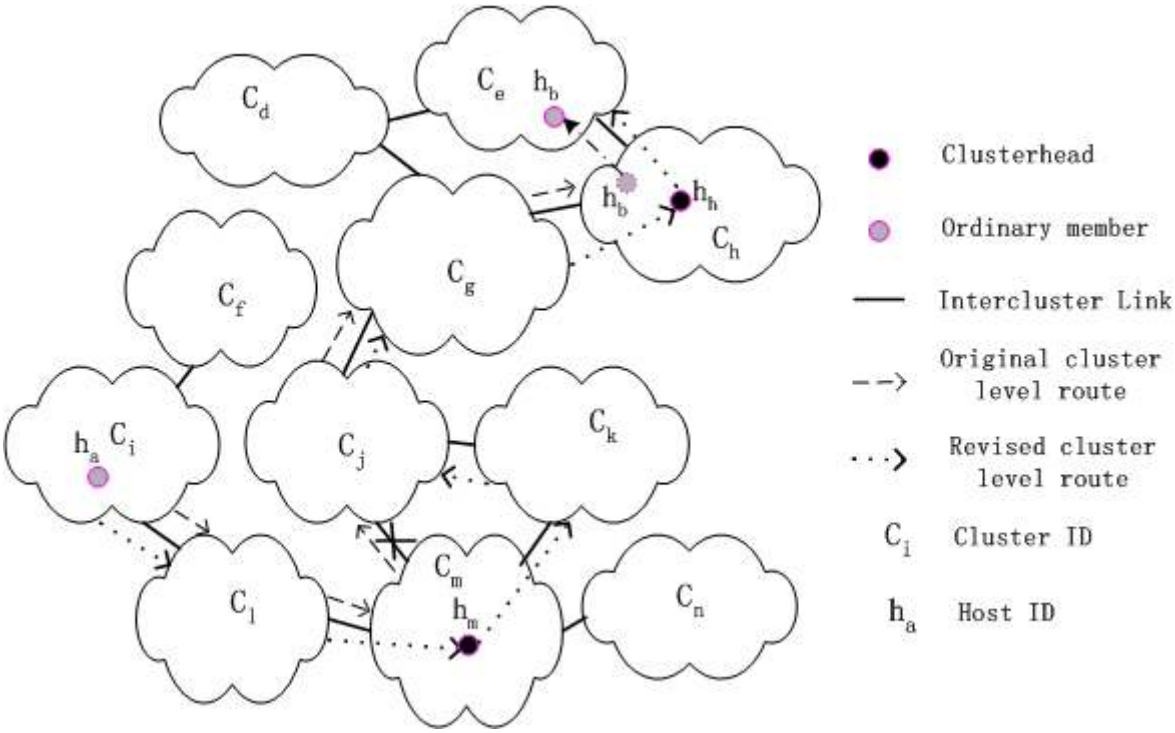


Figure 4-4. Cluster level routing and route recovery

Consequently, the KCBL protocol can enhance the performance of routing protocols. The high ability to tolerate link breakage and well scalability makes it competent for large scale networks. In the next chapter, the performance of KCBL is evaluated in a large scale ad hoc network.

4.3 Performance Analysis

Because the large size memory is available in personal computer nowadays, the storage capacity is no longer the obstacle to implement network protocols. Contrarily, control overheads are more critical due to the scarce wireless link bandwidth in ad hoc networks.

Therefore, this section first focuses on the overheads and cost caused by KCBL. The control overheads of KCBL are evaluated in the initial stage and the location maintenance stage, respectively. Secondly, the accuracy of location information provided by KCBL is discussed in an ad hoc network with variable host mobility.

Table 4-4. Simulation parameters for location service protocol

Items	Value
The number of hosts N	1000~4000
Maximum moving rate V_m	0.2~1 unit/s
Mean connection arrival rate λ_{Call}	1/420s
Scaling factor σ	0.1~0.5
Time interval T_{HM} [6]	10s
Distance threshold D_{th}	2 units

Simulation experiments are conducted in a homogeneous network where N mobile hosts are randomly distributed in the area with $S = 30 \times 30$ square units. The radio transmission range of each host is $r = 1$ unit. Every host moves according to a uniformly distributed moving rate among predefined ranges $[0, V_m]$, where V_m is the maximum host moving rate. The average connection arrival rate for each host is Poisson distributed with a mean value λ_{Call}. The parameters used in simulations are listed in Table 4-4. As the KCMBC scheme is able to generate more stable cluster structure than other cluster formation heuristics, the performance evaluation here is based on the clusters created by KCMBC. Each simulation runs 1000 seconds of simulation time. Moreover, the following notations will be used for performance evaluations.

- d_C : the average number of neighboring clusters around one cluster
- N_h : the average number of members in a cluster

[6] T_{HM} is the time interval which is used to calculation the average linke available time in a cluster.

- N_C : the total number of clusters in the network
- H_C : the average number of hops for an arbitrary intra-cluster route

4.3.1 Overheads in the Initial Stage

In the initial stage for the construction of network topologies, once clusters have been constructed, each cluster-head broadcasts a CS packet to all the other cluster-heads in the network. Upon receiving a CS packet, the gateway forwards it to the cluster-head. The cluster-head in turn selects the gateways to multicast this packet to other neighboring clusters. Therefore, each CS packet causes $d_C H_C N_C$ times transmission in the whole network. If H_C is approximated as k, and $N_C = N / N_h$, then the overheads caused by the CS packet transmissions for of all clusters is given by:

$$O_{CS} = d_C H_C N_C^2 \approx \frac{d_C k N^2}{N_h^2} \text{ packets}$$

Moreover, the overheads generated in the initial stage also include the overheads for cluster formation in the whole network. According to equation (29) in Chapter 3, under KCMBC, the total overhead for cluster formation is given by

$$O_{CF\_KCMBC} = (2k+3)N \text{ packets}$$

Consequently, the total overheads created in the initial stage can be obtained as

$$O_{Initial\_KCBL} = O_{CF\_KCMBC} + O_{CS} \approx (2k+3)N + \frac{d_C k N^2}{N_h^2} \text{ packets}, \quad (32)$$

where the values of d_C and N_h for KCMBC which are obtained by simulation can be found in Appendix D.

A discrete-event simulator is developed to count the amounts of communication overheads created by the KCBL protocol. Figure 4-5 shows both the simulation results and the numerical results predicted by equation (32). The number of hosts in the network varies from 1000 to 4000. From Figure 4-5, it can be observed that the simulation results match the predicted results. Moreover, it can be seen that overheads drastically decrease while the value of k increases. This is because that the increase of k leads to the increases of cluster size and the decrease of the number of clusters in the network. On the other hand, when k is a fixed value, $O_{Initial\_KCBL}$ increases with the number of hosts N. However, a large k suppresses the increase rate of $O_{Initial\_KCBL}$ when N increases. When $k = 4$, the difference among the overheads for variable N becomes little.

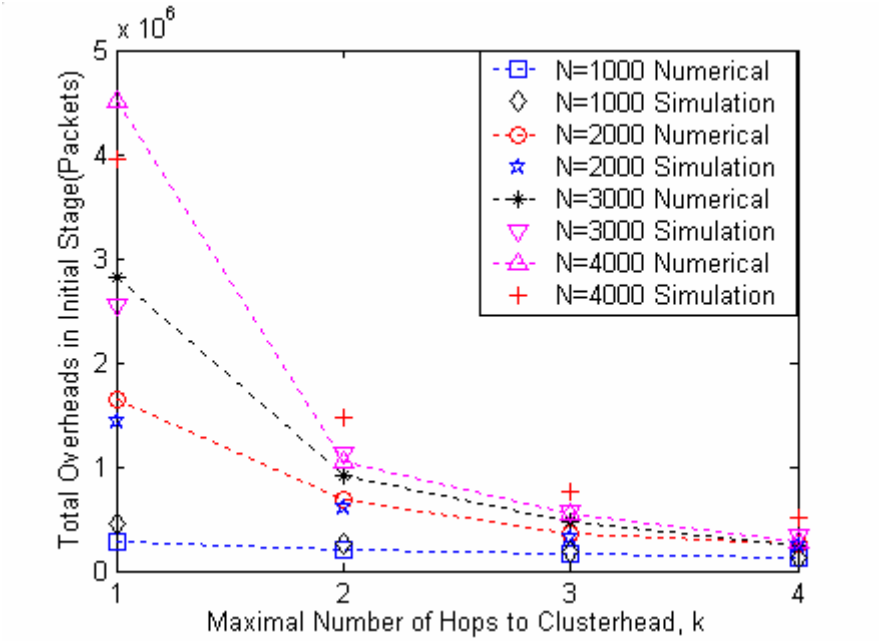


Figure 4-5. The overheads created by KCBL in the initial stage

The KCBL protocol can be considered as a hierarchical link state protocol: the global inter-cluster link states are maintained by cluster-heads, and the local intra-cluster link states are stored by each cluster member. It is interesting to compare the communication overheads in

the Link State Routing (LSR) protocol [82][90] with the overheads in the KCBL protocol. The LSR protocol is a typical flat proactive protocol, which applies Open Shortest Path First (OSFP) protocol [90]. In LSR, each host must maintain all link states in the network. As every host generates one link state packet which will be forwarded by all the other hosts once, the overheads generated by LSR in the initial stage are given by $O_{Initial\_LSR} = N^2$ packets.

Simulations for the LSR protocol validates that the predicted values $O_{Initial\_LSR}$ are correct. Table 4-5 shows the simulation results of $O_{Initial\_LSR}$. From the comparison of Figure 4-5 and Table 4-5, it can be found that the overheads in the initial stage by using KCBL protocol are much fewer than those by using the LSR protocol. For example, when $N = 4000$, $O_{Initial\_LSR}$ is more than 30 times of $O_{Initial\_KCBL}$ for $K = 4$.

Table 4-5. The overheads created by LSR in the initial stage

N	1000	2000	3000	4000
$O_{Initial\_LSR}$ (Packets)	1.00×10^6	4.00×10^6	9.00×10^6	16.00×10^6

4.3.2 Cost in Location Maintenance Stage

Considering that beacons are used in almost all routing protocols and location service protocols for ad hoc networks, the overheads of periodical beacons is not discussed here. Then in the location maintenance stage, the overheads created by the KCBL protocol mainly consist of the overheads for intra-cluster location updates and the overheads for inter-cluster location updates. The overheads for intra-cluster location updates are caused by intra-cluster link activation or deactivation. Any change of intra-cluster links will trigger a LLC packet, which must be delivered once by each cluster member. On the other hand, those link changes, which can vary cluster membership, cluster structure (such as cluster merge, cluster removal

and cluster reelection) or the inter-cluster connectivity, will trigger CS packets to update location information. In addition, a small amount of overheads are induced by location enquiries. The cost of location management, in units of control packets per second, is a summation of the cost of inter-cluster location update, C_{inter}, the cost of intra-cluster location update, C_{intra}, and the cost of location enquiry, C_{enq}. Consequently, the cost of location management is given by:

$$C_{KCBL} = C_{intra} + C_{inter} + C_{enq} \tag{33}$$

Table 4-6. The cost of location management

$V_m = 0.5$ Unit/sec, $\sigma = 0.2$

N / k Hops	1000			2000			3000			4000		
	C_{intra}	C_{inter}	C_{enq}	C_{intra}	C_{inter}	C_{enq}	C_{intra}	C_{inter}	C_{enq}	C_{intra}	C_{inter}	C_{enq}
1	2142	177078	5	10959	979494	10	35352	2289724	14	71641	3906724	19
2	3144	141807	10	23762	540904	19	83942	961523	29	205223	1655607	38
3	3920	102406	14	40573	358747	29	144170	504946	43	339582	617488	57
4	4752	74019	19	56220	249903	38	239750	250278	57	569116	443916	76

Table 4-6 lists the simulation results of C_{inter}, C_{intra} and C_{enq} with different values of k and N. It is obvious that the total cost is mainly dominated by C_{inter} and C_{intra}. The small cost induced by location enquiry can be ignored, since every location enquiry packet is transmitted within one cluster. From Table 4-6, it can be seen for a fixed value of N, when k increases, C_{intra} increases, whereas C_{inter} decreases. This is because with the increase of k, the average size of each cluster increases, and the total number of clusters in the network decreases accordingly. In this case, the number of transmissions for each LLC packet increases, and the number of clusters for forwarding a CS packet decreases. Table 4-6 also shows for a fixed k, the increase of the number of hosts leads to the increase of both C_{intra} and C_{inter}.

In the cost function (33), C_{inter} depends on the transmission interval of CS packets, which is determined by the scaling factor σ and link available time of each cluster. Figure 4-6 shows the effect of σ on the cost (C_{inter}) of inter-cluster location updates. It can be seen that for a fixed k, C_{inter} increases with the decrease of σ. This is because that a small σ results in small latency for transmitting CS packets among clusters, and accordingly, frequent inter-cluster location updates generate a large number of overheads. By contrast, a large σ can reduce the cost of inter-cluster location update. On the other hand, due to large transmission latency, a large σ will degrade the accuracy of location service provided by KCBL, which is described in Section 4.3.3.

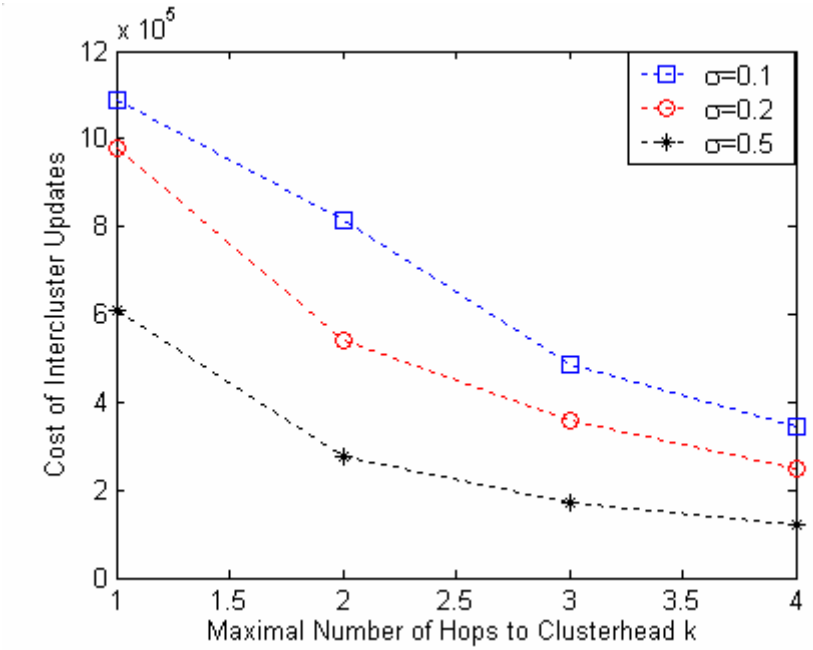


Figure 4-6. The cost of inter-cluster update with different scaling factor σ

$N = 2000$, $V_m = 0.5$ unit/sec

Figure 4-7 shows the total cost of location management C_{KCBL} versus k when $V_m = 0.5$ unit/sec. It is obvious that the total cost increases with the number of hosts N. However, a

large value of k can suppress the increase rate of C_{KCBL} while N increases. For example, when $k=1$, the total cost for $N=4000$ is about 22 times of that for $N=1000$; but when $k=3$, the total cost for $N=4000$ is only 9 times of that for $N=1000$.

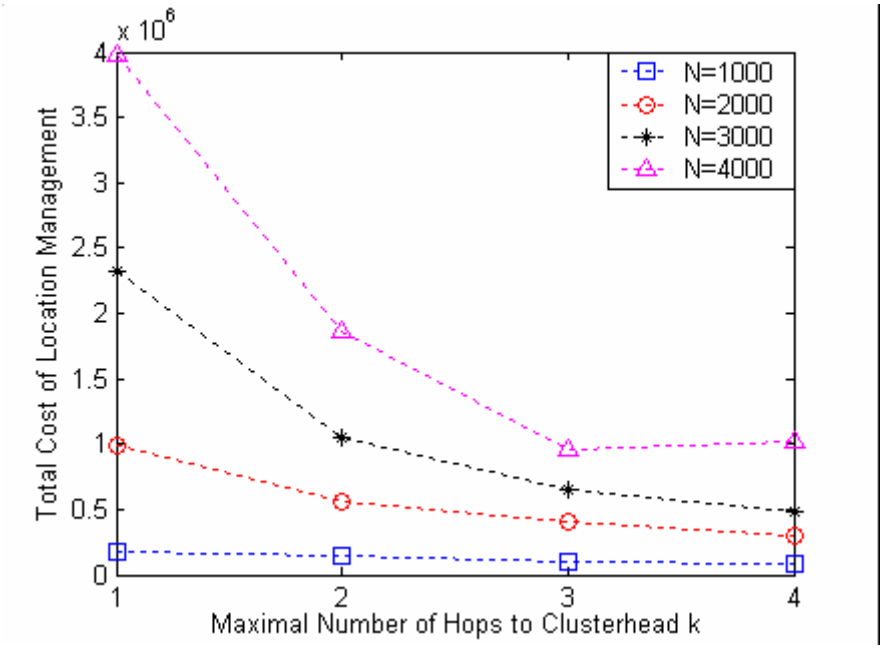


Figure 4-7. The total cost of location management C_{KCBL}

$V_m = 0.5$ unit/sec, $\sigma = 0.2$

Referring to Table 4-6, it can be found that the total cost is mainly determined by C_{inter} when the value of k is small. With the increase of k, C_{inter} decreases, whereas C_{intra} increases and prefers more contribution to the total cost than C_{inter}. Accordingly, as shown in Figure 4-7, when N has a fixed value, the total cost decreases with the increase of k. However, there is an exception when $N=4000$ and k changes from 3 to 4. From Figure 4-7, it also can be seen that the cluster structure with $k=1$ is not scalable for a large diameter ad hoc network. To reduce the total cost of location management, the value of k can be set as 3 or 4.

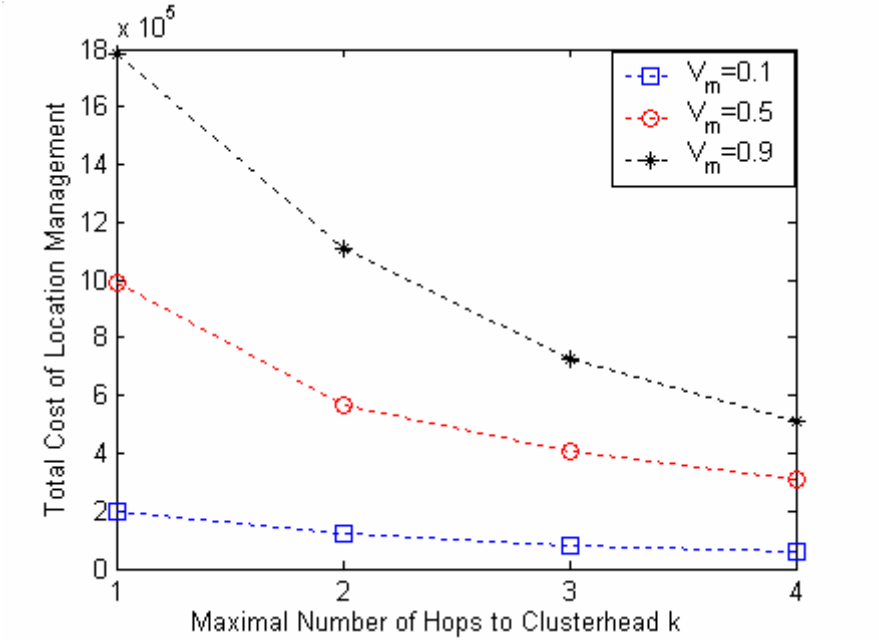


Figure 4-8 The effect of host mobility on C_{KCBL}.

$\sigma = 0.2$ and $N = 2000$

Figure 4-8 shows the effect of host mobility on C_{KCBL}. It is clear that C_{KCBL} increases with V_m, since both link states and cluster structures change frequently while host moving rate becomes large. However, a large k can suppress the increase rate of C_{KCBL} while the degree of host mobility increases. This is because that clusters with larger size can make a highly dynamic topology appear much less changing. From Figure 4-8, it can be found that cluster structures with a large k are able to reduce the cost of location management efficiently, especially in a high host mobility environment.

Furthermore, simulations are used to compare the cost of location management by using KCBL and the cost of link state updates in the LSR protocol [82][90], which are denoted as C_{KCBL} and C_{LSR}, respectively. Table 4-7 shows the cost of both protocols with different number of hosts in the network, when $V_m = 0.5$ unit/sec, $k = 3$ and $\sigma = 0.2$. It can be seen

that C_{LSR} increases with N following an approximately square rate, which is expected since any link change will trigger a location update packet that must be forwarded by every host once. From Table 4-7, it also can be found that C_{KCBL} is much less than C_{LSR} for a fixed N. For example, when $N = 4000$, C_{LSR} is more than 50 times of C_{KCBL}. The total cost of KCBL increases with N following a linear rate. The average cost for location management charged by each host is a sub-linear function of host density. For instance, the average cost of each host for $N = 4000$ is 2.26 times of that for $N = 1000$, whereas the host density for $N = 4000$ is 4 times of that for $N = 1000$. The simulation results support our claim that the proposed KCBL protocol is scalable to large and dense ad hoc networks.

Table 4-7. Cost comparison between LSR and KCBL

$V_m = 0.5$ unit/sec, $k = 3$ and $\sigma = 0.2$.

N	1000	2000	3000	4000
C_{LSR}	8.04×10^5	6.44×10^6	2.17×10^7	5.15×10^7
C_{KCBL}	1.06×10^5	3.99×10^5	6.49×10^5	9.57×10^5

4.3.3 Accuracy of Location Service

As described in Section 4.2.3, the KCBL protocol provides more accurate location information in the destination's neighborhood and less accurate information for the hosts far away. The accuracy of location service can be evaluated by the average hit probability of a location enquiry response. Specially, the hit probability $P_{HC}(n)$ is defined as the probability that the response of a location enquiry is able to provide the destination's current cluster ID correctly, where n denotes that n cluster-hop distance exists between the source-destination pair. In the simulations, $P_{HC}(n)$ is the average measured values of 5000 trials for each n.

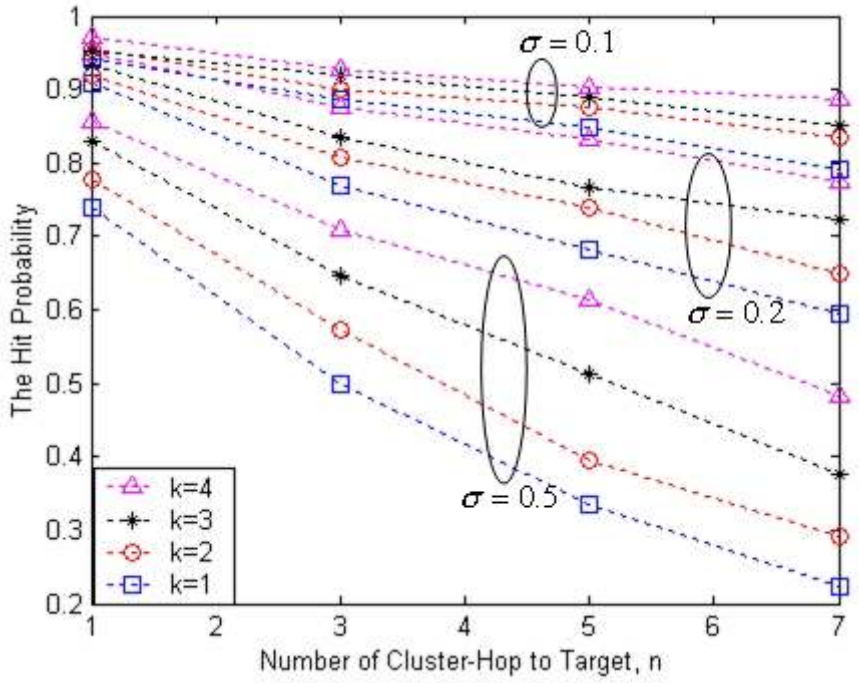
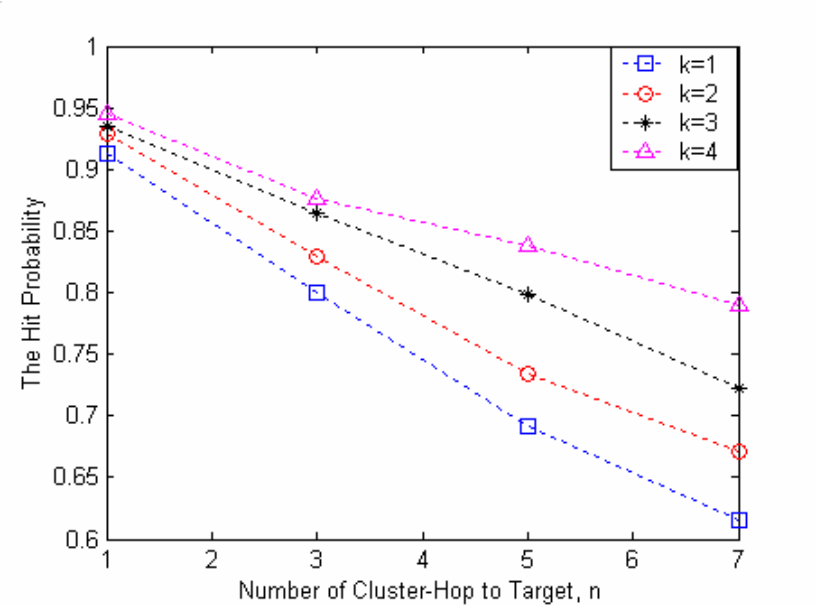


Figure 4-9. The hit probability versus cluster-hop distance n

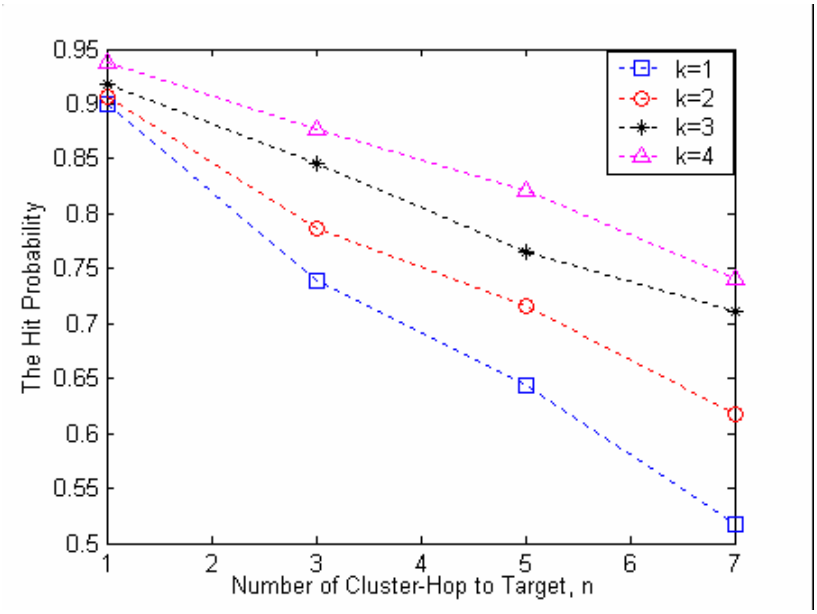
$N = 2000$, $V_m = 0.5$ unit/sec

Figure 4-9 illustrates the average hit probability versus cluster-hop distance n when $N = 2000$ and $V_m = 0.5$ unit/sec. It can be seen that $P_{HC}(n)$ increases with k, since the sojourn time of host in a cluster increases while the average cluster size increases. Figure 4-9 also shows that the average hit probability decreases with an approximately linear rate when the cluster-hop distance n increases. Since the latency of CS packet delivered to n cluster-hop distance away is $T_{Cn} = n\tau_{pi}$, the staler location information is obtained for the longer distance to the destination. For example, when $k = 2$ and $\sigma = 0.2$, the hit probability for $n = 7$ is about 65%. However, the hit probability for $n = 1$ is more than 90%. Therefore, stale location information can be revised by the intermediate cluster-heads which are closer to the destination.

From Figure 4-9, it also can be found that when n and k are fixed values, the hit probability decreases with the increase of the scaling factor σ. This is because that the transmission interval τ_{pi} for CS packets is positive proportional to σ. The latency for CS packet transmission increases with σ correspondingly. On the other hand, as discussed in Section 4.3.2, a small σ results in frequent inter-cluster location updates, and causes a large amount of control overheads. It is obvious that trade off exists between the cost of location management and the accuracy of location service. As the neighboring clusters around the destination maintains the latest location information, the value of σ can be determined by $P_{HC}(1)$, the average hit probability of a location enquiry response performed in the destination's direct neighboring clusters. In order to ensure $\Pr_{HC}(1)$ higher than 90% for any k, σ is set as 0.2 in most of our experiments. In this case, the cost of location management can be controlled at relatively small level (see Figure 4-7).



(a) $V_m = 0.1$ unit/sec



(b) $V_m = 0.9$ unit/sec

Figure 4-10. Impact of host mobility on the hit probability $P_{HC}(n)$, $N = 2000$

Figure 4-10(a) and (b) show the hit probability $P_{HC}(n)$ when the maximal host moving rate is 0.1 unit/sec and 0.9 unit/sec, respectively. It can be found that for both the moving scenarios, the hit probability increases with k, but decreases with the increase of n. Moreover, when n has a fixed value, there is a little difference (less than 10%) between $P_{HC}(n)$ in high host moving rate scenario ($V_m = 0.9$ unit/sec) and $P_{HC}(n)$ in low host moving rate scenario ($V_m = 0.1$ unit/sec). It is clear that the change of maximal host moving rate makes little influence on the hit probability. This is because in the KCBL protocol, the transmission interval for CS packets is proportional to the average link available time of corresponding clusters, which can reflect the degree of host mobility. The frequency of inter-cluster location updates increases with host moving rate. Therefore, the KCBL protocol is able to adapt the frequency of location updates in an ad hoc network with variable host mobility, and well balance the trade off between control overheads and the accuracy of location service.

4.4 Conclusion

In this chapter, we propose an efficient location service protocol, namely k -hop Clustering Based Location Service (KCBL) protocol. It utilizes the advantage of cluster architecture to construct a distributed location service system, where cluster-heads keep the location information of all clusters in the network and immediately respond to the location enquiries from their cluster members. The high stability of the KCBL protocol relies on the following reasons: (1) the influence of any inter-cluster link failure or cluster-head deactivation is limited within one cluster; (2) more than one physical link existing between two neighboring clusters makes the redundant cluster architecture tolerate false route and balance traffic load; (3) any stale location information can be updated along the path to the destination; (4) routing is still operated in a flat network structure, and cluster-heads do not risk becoming the communication bottleneck.

Location management of the KCBL protocol is performed on intra-cluster level and inter-cluster level. Considering the distance effect [54] and host mobility, an efficient inter-cluster location update mechanism is used to reduce control overheads significantly. KCBL provides more accurate location information in the destination's neighborhood and less accurate information for the hosts far away. The frequency of inter-cluster location updates can be determined by the mobility pattern of this cluster or the group characteristics. All of these above features ensure that KCBL has good scalability. In addition, due to the benefits of the cluster structure and distributed cluster-heads, KCBL is able to increase the performance of routing protocols, in terms

of overhead reduction and route recovery. The high ability to tolerate link breakage and good reliability makes KCBL efficient for large scale networks.

Simulation results indicate that both the overheads in the initial stage and the total cost in the location maintenance stage decrease with the increase of k. A large value of k can not only suppress the increase rate of the total cost while the number of hosts in the network increases, but also increase the hit probability of location service, and reduce the passive effect of host mobility on control overhead and the hit probability. Moreover, a suitable value of scaling factor σ can balance the trade off between the cost and the accuracy of location service. The simulation results also show that KCBL is scalable to various host density and host moving rate. With optimally chosen k and σ, the total cost of location management by using KCBL is only about of two percent of the cost of a link state protocol. Because of its good scalability and the ability of self-organization, the KCBL protocol can be used to provide location service for most routing protocols.

Chapter 5 Forwarding Host Selection Broadcast Relay Scheme

5.1 Introduction

Because of their ever-changing topology, ad hoc networks use broadcasting as one of the fundamental protocols. Especially when packets are transferred to multiple hosts, broadcast is more efficient than unicast or multicast. Blind flooding is the traditional broadcast approach for wireless networks. However, blind flooding generates a large number of redundant packets that waste valuable resources such as bandwidth and energy. The large number of redundant broadcast packets caused by blind flooding has been known as the Broadcast Storm Problem [92] in ad hoc networks. Current research on optimal broadcasting mechanisms in ad hoc networks has been focusing on minimization of rebroadcasts number and efficiency of delivering packets to hosts in the network. The routing protocols being currently designed for ad hoc networks can be divided into two categories, i.e. topology-based protocol and geometry-based protocol.

In the topology-based broadcast protocols [93][94][95][42], a mobile host determines where the received packet is forwarded to based on the connectivity information between the host and neighborhood host. Since optimal broadcasting forwarding in an ad hoc network has been proven as NP-complete [43], the topology-based protocols attempt to approximate minimal connected cover set. Each host is assumed to know the local connectivity information up to 2 hops. Based on the topology information, the topology-based protocols can generate a small set of neighborhood host for forwarding packets, so that the redundant

rebroadcasts can be significantly reduced while the maximum broadcast reach-ability is maintained. However, the exchange of local topology information under such topology-based protocol may cause a large amount of overhead, since each host must maintain a long list of the neighboring hosts, especially when the host density in the network is high. On the other hand, hosts in the network [43] [95][42] need to take long time for the collection of all connectivity information with the other neighboring hosts within 2 hops. During such long convergence time, the existing hosts may leave and the new hosts may join the 2-hop neighborhood range, which make the created forwarding host set incorrect. Hence the topology-based broadcast protocols cannot keep good performance in a highly dynamic environment.

By contrast, the geometry-based broadcast protocols choose the forwarding hosts according to the geometry location information of direct neighboring hosts. Each host obtains its location information through GPS receivers or some other type of position service [33][34]. The location information is exchanged among direct neighboring hosts via periodical beacons or broadcasting packets. The exchange of location information only consumes a small amount of bandwidth. Compared with topology-based protocols, the geometry-based protocols usually have less convergence time to obtain geometry location information, and they bring each host less computational load to generate the cover set. Therefore, the geometry-based protocols are more efficient for an ad hoc network with drastic host mobility. On the other hand, because of insufficient network topology information, the performance of geometry-based protocols is usually poorer than that of topology-based protocols.

Geometry-based broadcasting protocols including the distance-based protocol and the location-based protocol are presented in [49][92]. Under the distance-based protocol, the distance between the neighboring hosts that the broadcast has previously been communicated. When a new packet is received by a host, a random delay is initiated and redundant packets are cached. When the random delay expires, if and only if all hosts in the preceding hop are farther than a threshold distance value, the host rebroadcasts the packet. By contrast, under the location-based approach [92], the hosts receiving packets calculate the additional coverage that can be offered using the location information of the transmitting hosts located in the preceding hop. A predefined threshold area value is used to determine whether the receiving host should rebroadcast when the random delay expires. Variable threshold defined as the function of the host density are studied in [49]. It has been demonstrated that both of the distance-based approach and the location-based approach are able to reduce the number of redundant broadcasts. However, these two protocols may also incorrectly drop important rebroadcasts, which result in low reach-ability, especially when the network has sparse host density.

In order to achieve the maximal reach-ability, the angle-based protocol proposed in [50] uses the cover angles of forwarding hosts to determine the rebroadcast. When a new packet is received by host X, the host X initiates a random delay. After the random delay expires, the host X rebroadcast the received packet only if the rebroadcast of the same packet from the other neighboring hosts does not fully cover X's transmission range. However, in order to reach a high reach-ability, the angle-based approach generates large number of massive redundant rebroadcasts.

This chapter presents a novel Forwarding Host Selection (FHS) broadcast relay scheme to improve the performance of geometry-based protocols in terms of increasing broadcast coverage area. The broadcast scheme with high broadcast coverage area can achieve high delivery rate with low number of rebroadcasts. Firstly, the analysis of the coverage area for both single-hop broadcast relay case and multihop broadcast relay case are presented, in which each broadcasting host is assumed to have the same number of neighboring hosts taking part in the rebroadcasts. The upper bound of broadcast coverage is obtained, which is used for the design of the FHS scheme. Differing from common geometry-based schemes, the FHS scheme is based on both relative distance and forward angle information of neighboring hosts. With little extra overhead and computational load, FHS can help many geometry-based broadcast protocols [49][50][52][92] to achieve high delivery rate and low number of rebroadcasts.

5.2 The Upper Bond of Broadcast Coverage Area

This section focuses on the upper bound of broadcast coverage area under the condition that each broadcasting host is assumed to have the same number of neighboring hosts which involve the rebroadcasts. An ad hoc network is modeled using Unit Disk Graph [19]. As shown in Figure 5-1, each host has a transmission range of radius $r = 1$ unit. Two hosts X and Y in the network are neighboring hosts with the Euclidean distance $|XY| = x \le 1$. The forwarding host is defined as the host which involves the broadcast relay process, and the neighboring host, which rebroadcasts the received packet, is named as forwarding neighboring host. Without loss of generality, let $TA_{(k,n)}$ denote the total coverage area of $m(k,n)$ broadcasting hosts, and $AS_{(k,n)}$ be the average coverage

area of each broadcasting host in the area, where k represents the number of rebroadcast hops from the source host, n presents the number of broadcasting neighboring hosts around each forwarding host, and $m(k,n)$ represents the total number of broadcasting hosts in the k-hop broadcast. The following analyzes the coverage area of both single-hop broadcast relay case and multihop broadcast relay case.

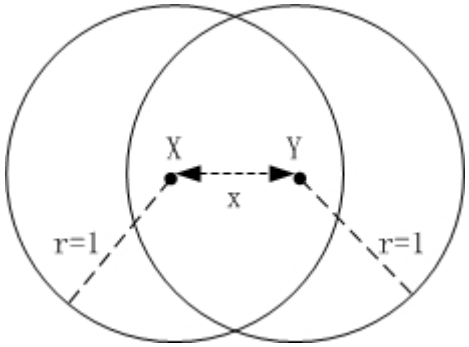


Figure 5-1. The coverage area of two neighboring hosts

5.2.1 Single Hop Broadcast Relay

Figure 5-1 shows a single hop broadcast relay case which consists of two neighboring forwarding hosts X and Y. In this case, the broadcast coverage area only depends on the distance $|XY| = x$ between the two hosts. Let $TA_{(1,1)}(x)$ denote the radio coverage area of the two neighboring hosts, then $TA_{(1,1)}(x)$ can be given by

$$TA_{(1,1)}(x) = \pi + 2\arcsin\frac{x}{2} + x\sqrt{1-(\frac{x}{2})^2}, \quad (0 \le x \le 1)$$

The differentiation equation of $TA_{(1,1)}(x)$ is obtained as

$$\frac{dTA_{(1,1)}(x)}{dx} = 2\sqrt{1-(\frac{x}{2})^2} > 0, \quad (0 \le x \le 1)$$

Thus, when $x = 1$, $TA_{(1,1)}(x)$ can reach the maximum value, that is

$$Max(TA_{(1,1)}) = \frac{4\pi}{3} + \frac{\sqrt{3}}{2} \approx 1.609\pi$$

In this case, hosts X and Y are located on the transmission range of each other. The maximum value of the average coverage area of each forwarding host is given by

$$Max(AS_{(1,1)}) = Max(TA_{(1,1)})/2 \approx 0.804\pi \tag{34}$$

Figure 5-2 illustrates the case of single-hop broadcast relay consisting of the source X and n forwarding hosts Y_i (i=1,2,$\cdots$,n), which take part in the rebroadcasts among X's neighboring hosts. Besides the distance $|XY_i|$, the forward angles α_i between X and Y_i have effects on the total coverage range. Theorem 1 indicates the condition of the forward angles to obtain the maximum value of the average broadcast coverage area.

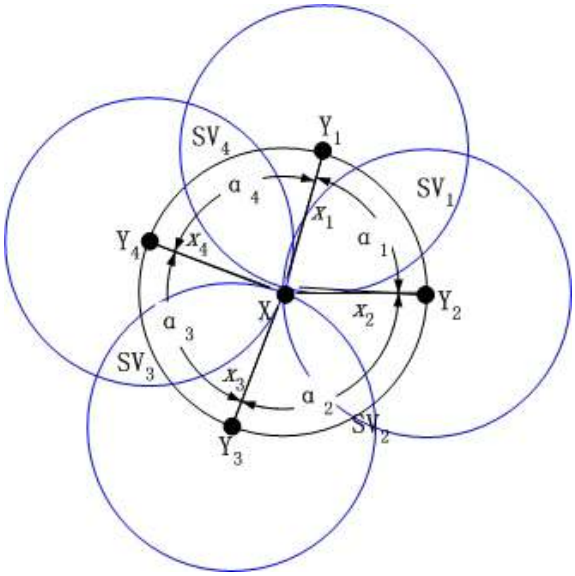


Figure 5-2. Single-hop broadcast relay, $n = 4$

Theorem 1. If the source X has n ($n \geq 3$) broadcasting neighboring hosts Y_i (i=1,2,$\cdots$,n), the total broadcast coverage area $TA_{(1,n)}$ can reach the maximum value

when all forwarding neighboring hosts Y_i are symmetrically located on the border of X 's transmission range.

Proof. From the single hop broadcast relay case with two forwarding neighboring hosts, it can be deduced that the total broadcast coverage area for $|XY_i| = x_i = 1$ is larger than that for $x_i < 1$.Referring to Figure 5-2, when $x_i = 1$, the total broadcast coverage area is given by

$$TA_{(1,n)} = \pi + n*(\frac{\pi}{3} + \frac{\sqrt{3}}{2}) - \sum_{i=1}^{n} SV_i(\alpha_i)$$

and $\sum_{i=1}^{n} \alpha_i = 2\pi$,

where α_i is the forward angle between X and Y_i , $\alpha_i = \angle Y_i X Y_j, j = (i+1) \ mod \ n$, and SV_i is the overlapping area caused by the transmission ranges of two adjacent forwarding neighboring hosts. When $\alpha_i \geq 2\pi/3$, $SV_i(\alpha_i) = 0$.

It is obvious that the maximum value of $TA_{(1,n)}$ can be obtained when $\sum_{i=1}^{n} SV_i(\alpha_i)$ is minimum. By using Lagrangean relaxation technique [96], let

$$g(\alpha_1, \cdots, \alpha_n) = \sum_{i=1}^{n} \alpha_i - 2\pi$$

Define that $F = \sum_{i=1}^{n} SV_i(\alpha_i) + \beta g(\alpha_1, \cdots, \alpha_n) = \sum_{i=1}^{n} SV_i(\alpha_i) + \beta(\sum_{i=1}^{n} \alpha_i - 2\pi)$, (35)

where β is the Langrange multiplier.

The minimum value of F can be reached under the following condition:

$$\begin{cases} \partial F / \partial \alpha_1 = \partial SV_1(\alpha_1) / \partial \alpha_1 + \beta = 0 \\ \qquad\qquad \vdots \\ \partial F / \partial \alpha_n = \partial SV_n(\alpha_n) / \partial \alpha_n + \beta = 0 \\ \partial F / \partial \beta = \sum_{i=1}^{n} \alpha_i - 2\pi = 0 \end{cases} \tag{36}$$

Thus, $\partial SV_1(\alpha_1) / \partial \alpha_1 = \partial SV_2(\alpha_2) / \partial \alpha_2 = \cdots = \partial SV_n(\alpha_n) / \partial \alpha_n$,

where $\alpha_1, \cdots, \alpha_n$ are coordinative in (35), the only solution of Equation (36) is:

$\alpha_i = 2\pi / n \quad (i = 1, 2 \ldots n)$.

Therefore, $Max(TA_{(1,n)})$ can be obtained if and only if:

$$\begin{cases} \alpha_i = 2\pi / n \\ x_i = 1 \end{cases}, \quad i = 1, 2, \ldots n \tag{37}$$

In other words, $TA_{(1,n)}$ can reach the maximum value when all forwarding neighboring hosts Y_i are symmetrically located on the border of X 's transmission range.

Based on Theorem 1, the maximum value of the average area covered by each broadcasting host can be derived accordingly. For the single-hop broadcast relay case with $n = 3$, the maximum value of $TA_{(1,3)}$ is obtained when forwarding neighboring hosts Y_i $(i = 1, 2, 3)$ are located on the border of X 's transmission range with forward angle $\alpha_i = 2\pi / 3$, that is

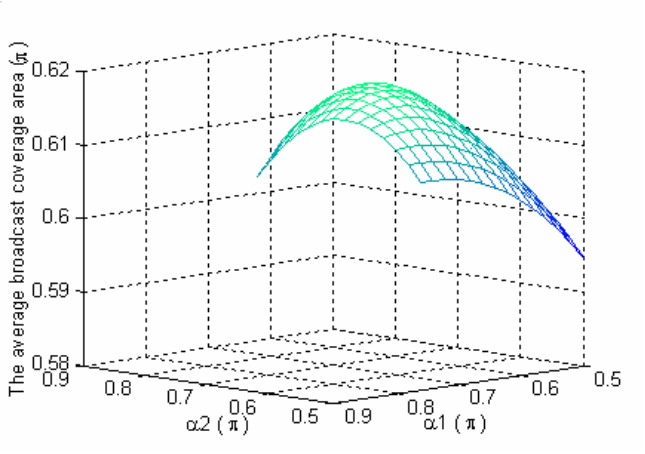
$$Max(TA_{(1,3)}) = \pi + 3 * (\frac{\pi}{3} + \frac{\sqrt{3}}{2}) \approx 2.827\pi \tag{38}$$

The maximum average coverage area of each forwarding host can be obtained as

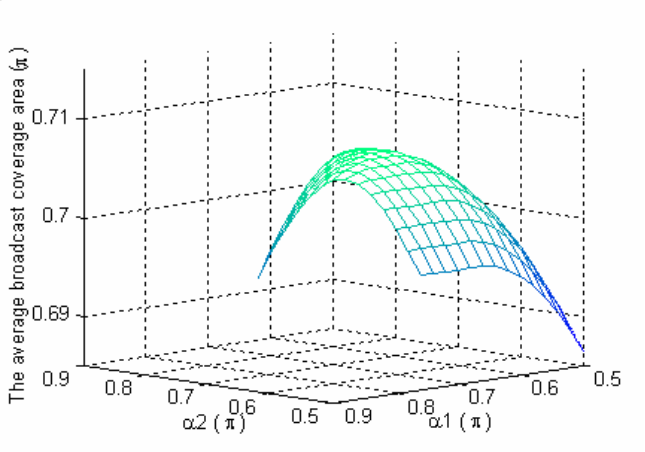
$$Max(AS_{(1,3)}) = Max(TA_{(1,3)}) / 4 = 0.706\pi \tag{39}$$

Figure 5-3 shows the average coverage area of each broadcasting host for the single-hop broadcast relay case when $n = 3$ and the distance between X and each Y_i $(i = 1, 2, 3)$ has the same value x. It is clear that $AS_{(1,3)}$ reaches the maximum value when $\alpha_i = 2\pi / 3$ $(i = 1, 2, 3)$, and $AS_{(1,3)}$ decreases when the deflection between the forward

angle α_i and $2\pi/3$ increases. From Figure 5-3, it can be found that when α_i have fixed values, the coverage area for $x=1$ is always larger than that for $x=0.8$.



(a) $x=0.8$



(b) $x=1$

Figure 5-3. $AS_{(1,3)}$ for the case of single-hop broadcast relay

$$x_i = x \ (i=1,2,3) \text{ and } \sum_{i=1}^{3} \alpha_1 = 2\pi$$

According to Theorem 1, if the number of forwarding neighboring hosts increased to $n=4$, the maximum value of $TA_{(1,4)}$ is obtained when forwarding neighboring hosts Y_i $(i=1,2,3,4)$ are located on the border of X's transmission range with forward angle $\alpha_i = \pi/2$. In this case, referring to Figure 5-2,

$$Min(\sum_{i=1}^{4} SV_i(\alpha_i)) = 4 * SV(\pi/2) = 4 * (\frac{\sqrt{3}}{2} + \frac{\pi}{12} - 1)$$

The maximum value of $TA_{(1,4)}$ is given by

$$Max(TA_{(1,4)}) = \pi + 4 * (\frac{\pi}{3} + \frac{\sqrt{3}}{2}) - min(\sum_{i=1}^{4} SV_i(\alpha_i)) = 2\pi + 4 \approx 3.273\pi$$

The maximum value of the average coverage area of each forwarding host can be obtained as

$$Max(AS_{(1,4)}) = Max(TA_{(1,4)})/5 = 0.655\pi \quad (40)$$

From the comparison of equation (34), (39) and (40), it can be found that the maximum value of $AS_{(1,n)}$ decreases when n increases. This is because that the overlapping area of neighboring hosts increases with the value of n. However, when $n < 3$, the broadcast generates coverage interstice which significantly degrades broadcast reach-ability. Hence the average coverage area of each forwarding host reaches the upper bound when Y_i (i=1,2,3) are located around host X with $\alpha_i = 2\pi/3$ and $x_i = 1$. The next subsection hereby only focuses on the multihop broadcast relay cases for $n = 3$ and $n = 4$ broadcasting neighboring hosts around each forwarding host.

5.2.2 Multihop Broadcast Relay

Figure 5-4 shows the multihop broadcast relay case for $n = 3$ broadcasting neighboring hosts around each forwarding host. In this case, the maximum broadcast coverage area can be obtained when (1) two forwarding neighboring hosts Z_j (j=1,2,…,6) are located on the border of Y_i's transmission range, and (2) all triangles between X, Z_g

and Z_{g+1} (g=1,3,5) are equilateral triangles. As shown in Figure 5-4, let S_z denote the area overlapped by two adjacent hosts Z_j and Z_m, where $m=(j+1)\ mod\ 6$. S_Z can be obtained as

$$S_Z = 2\times\int_0^1 (\sqrt{1-(y-\frac{1}{2})^2}-\frac{\sqrt{3}}{2})dy = \frac{\pi}{3}-\frac{\sqrt{3}}{2}.$$

The maximum value of $TA_{(2,3)}$ can be given by

$$Max(TA_{(2,3)}) = Max(TA_{(1,3)}) + 6\times(\frac{\pi}{3}+\frac{\sqrt{3}}{2}) - 3\times S_Z = Max(TA_{(1,3)}) + \pi + 9\sqrt{3}/2.$$

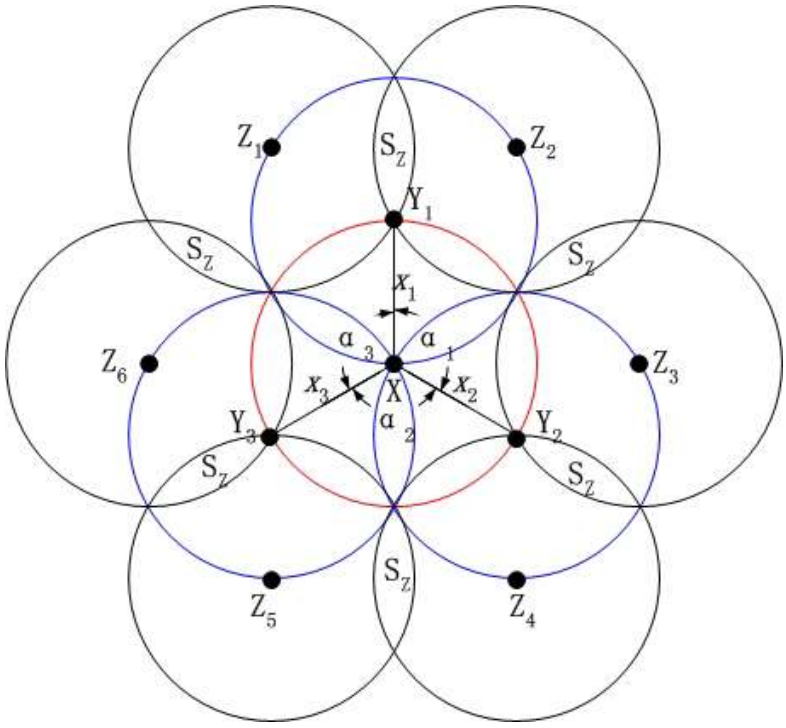


Figure 5-4. Multihop broadcast relay

3 forwarding neighboring hosts around each host $\alpha_i = 2\pi/3$

Moreover, when $k\geq 1$, $TA_{(k,3)}$ reaches the maximum value when all forwarding neighboring hosts are symmetrically located on the border of broadcasting hosts' transmission range, that is:

$$Max(TA_{(k,3)}) = \begin{cases} Max(TA_{(k-1,3)}) + \pi + (9k-3)\sqrt{3}/4, & k = 2t-1 \\ Max(TA_{(k-1,3)}) + \pi + 9k\sqrt{3}/4, & k = 2t \end{cases},$$

where t=1,2,3,…, and $Max(TA_{(0,3)}) = \pi$.

From the above iterative equation, $Max(TA_{(k,3)})$ can be given by

$$Max(TA_{(k,3)}) = \begin{cases} (k+1)\pi + (9k^2+6k-3)\sqrt{3}/8, & k = 2t-1 \\ (k+1)\pi + (9k^2+6k)\sqrt{3}/8, & k = 2t \end{cases} \quad (41)$$

The total number of forwarding hosts is given by

$$m_{(k,3)} = 1+3+3*2+\cdots+3k = \frac{3k^2+3k+2}{2}, \quad k \geq 1$$

Hence, the maximum average coverage area of each forwarding host can be obtained as

$$Max(AS_{(k,3)}) = \frac{Max(TA_{(k,3)})}{m_{(k,3)}} = \begin{cases} \dfrac{2(k+1)\pi + (9k^2+6k-3)\sqrt{3}/4}{3k^2+3k+2}, & k = 2t-1 \\ \dfrac{2(k+1)\pi + (9k^2+6k)\sqrt{3}/4}{3k^2+3k+2}, & k = 2t \end{cases} \quad (42)$$

Likewise, for $n = 4$, the maximum broadcast coverage area can be obtained by

$$Max(TA_{(k,4)}) = Max(TA_{(1,4)}) + \sum_{w=2}^{k}(\pi + 4w) = 2k^2 + (\pi+2)k + \pi \quad (43)$$

The total number of forwarding hosts is given by

$$m_{(k,4)} = 1+4+4*2+\cdots+4k = 2k^2+2k+1$$

The maximum value of $AS_{(k,4)}$ can be expressed by

$$Max(AS_{(k,4)}) = \frac{Max(TA_{(k,4)})}{m_{(k,4)}} = \frac{2k^2+(\pi+2)k+\pi}{2k^2+2k+1} \quad (44)$$

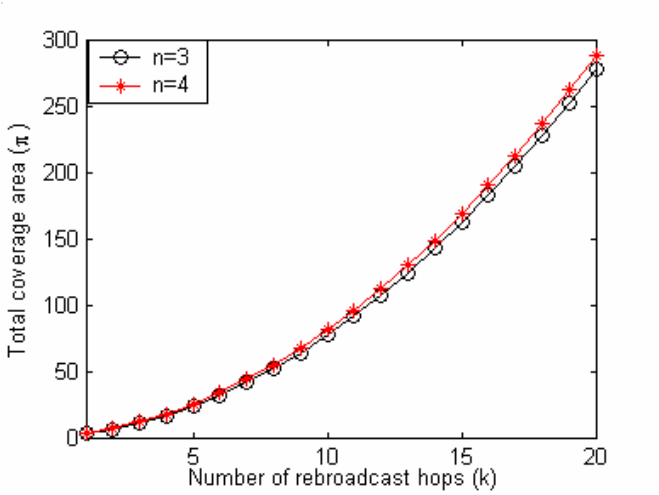


Figure 5-5. $Max(TA_{(k,n)})$ vs. the number of rebroadcast hops

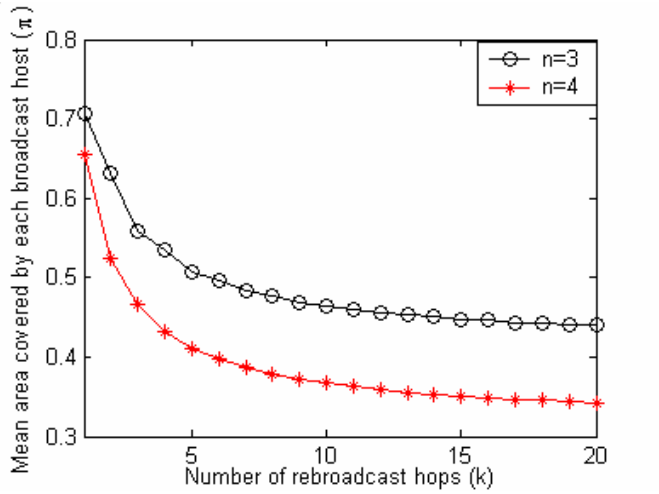


Figure 5-6. $Max(AS_{(k,n)})$ vs. the number of rebroadcast hops

Figure 5-5 shows the maximum value of total coverage area versus the number of rebroadcast hops, for $n=3$ and $n=4$. It can be seen that the value of $Max(AS_{(k,n)})$ increases when the number of broadcast hops increases. The values of $Max(TA_{(k,n)})$ for $n=3$ and for $n=4$ are always similar. From Figure 5-6, it can be found for fixed k, $Max(AS_{(k,n)})$ decreases when n increases. This is because that the overlapping area of neighboring hosts increases with n. Consequently, when forwarding hosts are

symmetrically located with $n = 3$, the average area reaches the upper bound which is determined by equation (42).

Although mobile hosts in an ad hoc network are randomly distributed and all forwarding hosts are hardly located on the ideal symmetrical position, this upper bound of broadcast coverage can be reached by moving each host to the ideal symmetrical position for the applications such as ad hoc sensor networks [10]. On the other hand, this obtained upper bound is useful for the design of novel broadcast relay scheme to improve the performance of geometry-based broadcast protocols.

5.3 The FHS Scheme

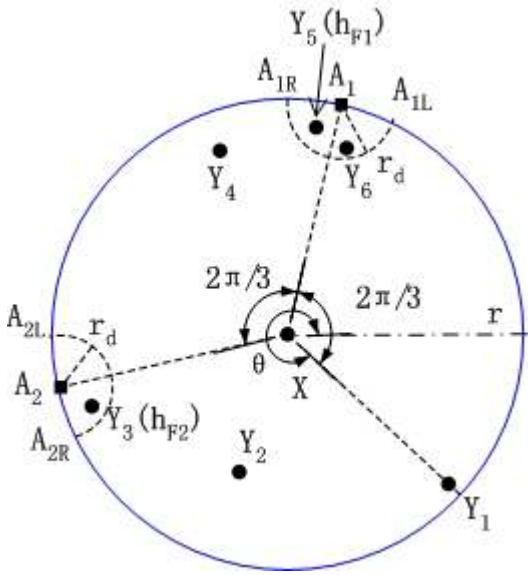


Figure 5-7. The fast forwarding host selection in FHS

To approach the upper bound of broadcast coverage, a novel FHS broadcast relay scheme is proposed. It is assumed that mobile hosts are randomly located in a two-dimensional homogeneous ad hoc network. Each host obtains its location information through GPS

receivers or some other type of position service [33][34]. Location information is exchanged by periodical "Hello" packets among neighboring hosts. Under the FHS scheme, each host that wishes to forward a packet calculates the coordinates of the symmetrical area around itself. The neighboring hosts located inside of the symmetrical area will be selected as fast forwarding hosts, which have higher priority to rebroadcast than other forwarding hosts. As shown in Figure 5-7, host X that wishes to broadcast a packet will select fast forwarding hosts by using the following algorithm.

Algorithm for selecting fast forwarding hosts

Input: the locations of n hosts inside of the transmission range of host X,

the reference direction θ

Output: one or two fast forwarding hosts

1) Host X calculates the planar coordinates of the symmetrical points A_i (shown in Figure 5-7) around itself using

$$A_i: (X_x + cos(\theta + 2i\pi/3), X_y + sin(\theta + 2i\pi/3)), \quad i = 1,2, \tag{45}$$

where (X_x, X_y) denote X's coordinates.

2) For each neighboring host Y_j ($j = 1,2,\ldots n$), the host X detects if $|Y_j A_i| \le r_d$, where $i = 1,2$, and r_d is a predefined threshold. Among the neighboring hosts who match the condition of $|Y_j A_i| \le r_d$, the nearest host to A_1 is selected as one fast forwarding host, denoted as h_{F1}, and the nearest host to A_2 is selected as another fast forwarding host, denoted as h_{F2}

When the source host X intends to broadcast a packet m, X selects a farthest neighboring host within its transmission range as one fast forwarding host, denoted as h_{F0}, and sets θ,

the relative direction from X to h_{F0}, as the reference direction. Through the above algorithm, the fast forwarding hosts h_{F1} and h_{F2} are selected among X's neighboring hosts which are located inside of the crescent areas $A_i A_{iL} A_{iR}$ $(i = 1,2)$ shown in Figure 5-7. For example, as shown in Figure 5-7, hosts Y_1, Y_3 and Y_5 are selected as the fast forwarding hosts h_{F0}, h_{F1} and h_{F2}, respectively. On the other hand, if an intermediate host X receives a packet m from the preceding host Y_1, when host X decides to rebroadcast packet m, it sets the reference direction θ as the relative direction from X to Y_1, .and selects the fast forwarding hosts h_{F1} and h_{F2} by using the above algorithm. Then, host X broadcasts packet m piggybacked with h_{Fi}'s IDs.

However, if there is no host located in the inside of the symmetrical areas $A_i A_{iL} A_{iR}$, h_{F1} or (and) h_{F2} is (are) not available, then the broadcasting coverage gaps can be filled by the rebroadcasts from the other neighboring hosts in the area.

5.3.1 The FHS Angle-Based Protocol

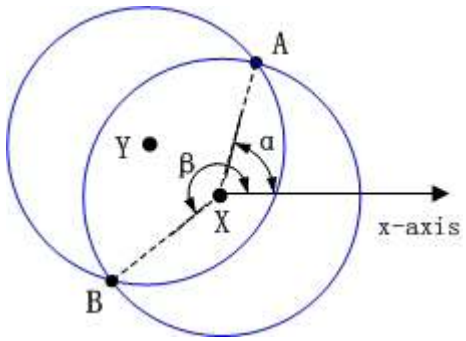


Figure 5-8. Cover angle

The proposed FHS scheme combined with the angle-based broadcast protocol [50], named as the FHS angle-based protocol, can be described as follows. Upon receiving a packet m from host Y, host X calculates the cover angle $\angle AXB$ (counterclockwise) as shown in Figure 5-8, where A and B are the intersections of the two circles centered at X and Y. Let the positive direction of x-axis be the reference direction of zero degree, then the directions of A and B relevant to host X can be expressed by the angles α and β, respectively. The interval $[\alpha, \beta]$ is referred to as the cover range of Y. It is supposed that during a transmission defter time interval, X has heard the rebroadcast of the same packet m from a number of neighboring hosts with cover range $[\alpha_1, \beta_1], \ldots, [\alpha_k, \beta_k]$. If the union of the cover ranges of the neighboring hosts is $[0, 2\pi]$, i.e., $\cup_j [\alpha_j, \beta_j] = [0, 2\pi]$, the whole transmission range of X must have been coved by the rebroadcasts from its neighboring hosts. In this case, host X will not further relay packet m.

Let n denote the number of host X's neighboring hosts. Upon receiving a packet m from one of its neighboring hosts, which is denoted as Y_j ($j = 1, 2, \ldots n$), host X executes the broadcast relay process as shown in Figure 5-9. In this process, packet m will be discard if at least one of the following conditions is satisfied: (a) the packet m has been forwarded before; (b) host X has only one neighboring host, i.e., $n = 1$; (c) the number of rebroadcast hops reaches to a predetermined time-to-live (TTL) value. If the packet m is not discarded and the host X is the fast forwarding host selected by the preceding hop, then the packet m must be relayed immediately by the host X; otherwise, the host X must wait a defer time. Since a neighboring host far from the preceding host Y_j usually covers more new area than a neighboring host near to Y_j, the defer time can make the farther neighboring host relay packets earlier. Using distance-based defer time in [50], the defer time is set as:

$$def(m) = MaxDeferTime \cdot (r^2 - \left|XY_j\right|^2 / r^2) \quad (46)$$

where r is the radius of the transmission range for a host. During the defer time, host X will discard packet m, if the duplicated copies of packet m are received and the union of the cover ranges of the neighboring hosts is $[0, 2\pi]$. The above procedure goes on until packet m reaches its destinations or the TTL value of packet m drops to zero.

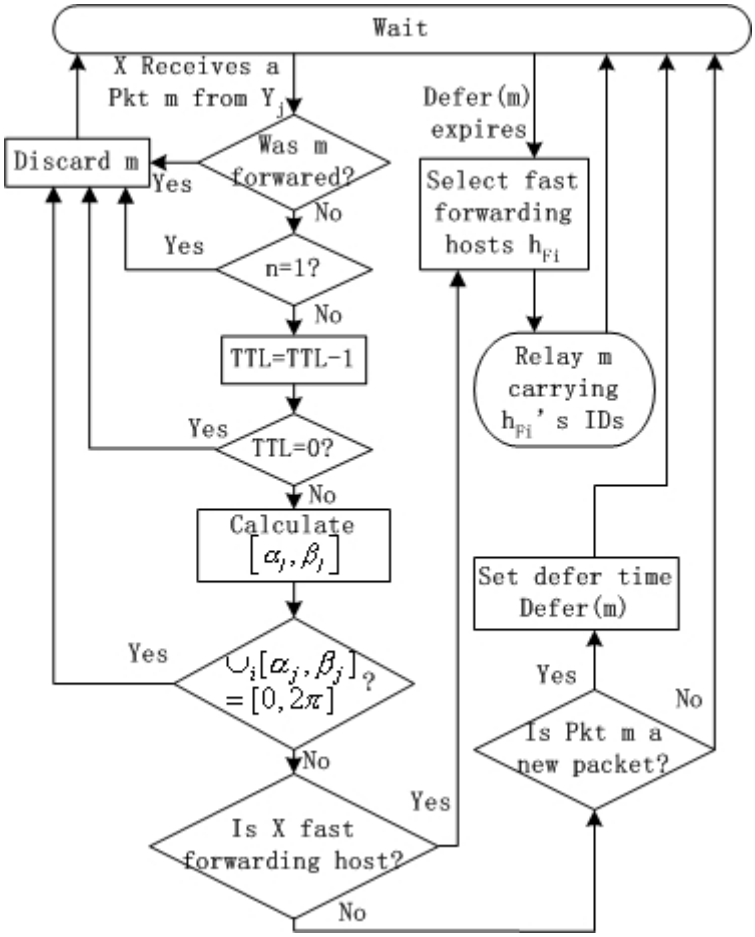


Figure 5-9. Flow chart of the FHS angle-based protocol

As the fast forwarding hosts are located inside of the symmetrical area, they are able to cover more new area. Considering a forwarding host X, the symmetrical locations of the two fast forwarding hosts and the preceding host Y_j make that the union of the three neighboring

hosts' angle cover ranges achieves a large value. Hence, only a small number of rebroadcasts from other neighboring hosts can fill up the gaps of angle cover ranges, and redundant rebroadcasts are reduced. On the other hand, it is obvious that the selected fast forwarding hosts have higher priority to deliver packets than other forwarding neighboring hosts. Therefore, the forwarded packets can be quickly propagated throughout the network, and the total transmission latency can be decreased. Another benefit of using FHS scheme is that the rebroadcasts with different priority are able to decrease medium access collisions.

5.3.2 The FHS Distance-Based Protocol

The FHS scheme is not only used in the angle-based protocol, but also can be used to improve the performance of other geometry-based broadcast protocols such as distance-based protocol, location-based protocol and counter-based protocol [49][92]. As another example, the proposed FHS scheme combined with the distance-based broadcast protocol [50], named as the FHS distance-based protocol, is described in this subsection.

Basically, the FHS distance-based protocol uses the same algorithm to select the fast forwarding hosts, which can rebroadcast packets without a defer time. By contrast, other hosts intend to relay packets must wait a distance-based defer time. The process skeleton of the FHS distance-based protocol is similar to that of the FHS angle-based protocol. However, it is noticed that under the distance-based protocol, only if all hosts in the preceding hop are farther than a threshold distance D_{th}, the host can be allowed to rebroadcast the packet. In an ad hoc network with sparse host density, it is possible that the distance between a forwarding host and each of its neighboring hosts is less than D_{th}. In this case, the broadcast process has to be terminated, and broadcast reach-ability may be largely degraded.

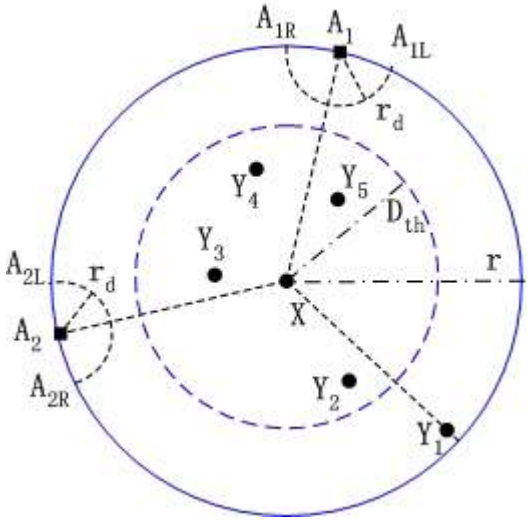


Figure 5-10. Forwarding host selection using the FHS distance-based protocol

The FHS distance-based protocol uses compensative forwarding host to solve the above problem. As shown in Figure 5-10, host X receives a packet m from the preceding hop Y_1. When host X intends to rebroadcast packet m, X calculates the relative positions of its neighboring hosts Y_j ($j = 2,\ldots n$). If no fast forwarding host is available and $|XY_j| < D_{th}$ for every j, host X selects the neighboring host, which has the longest distance to the preceding host Y_1, as the compensative forwarding host (for example, host Y_4 shown in Figure 5-10), denoted as h_C. Then, containing h_C's ID, packet m is sent out by X. The compensative forwarding host and the fast forwarding host can be distinguished by using different flags. Upon receiving packet m, host Y_j will immediately forward this packet, if packet m is heard for the first time and Y_j is selected as the compensative forwarding host h_C. Figure 5-11 shows the process of the FHS distance-based protocol.

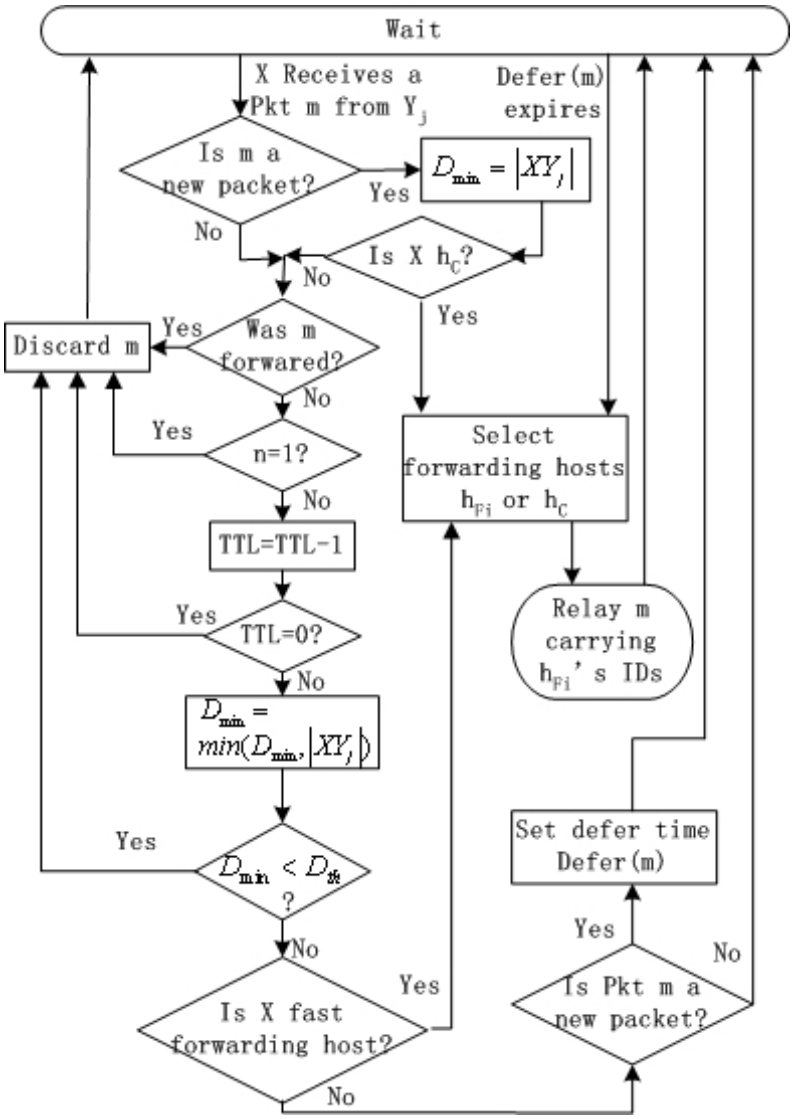


Figure 5-11. Flow chart of the FHS distance-based protocol

Moreover, an acknowledgment (ACK) mechanism can be used to combat collisions and high error rate in ad hoc networks. Upon receiving a correct packet, the selected forwarding hosts h_{Fi} or h_C will reply an ACK to the source X. On the other hand, once host X has broadcasted a packet m with the IDs of the available h_{Fi} or h_C, host X must wait for ACK packets from the selected forwarding hosts. If all of the relevant ACK packets cannot be received during a time period *ACKTimeout*, host X will retransmit packet

m . Through this broadcast handshake, most packets can be transmitted correctly even in a high error rate environment.

FHS is expected to be more effective in a network with high host density, because more forwarding hosts can be found within the symmetrical areas. With a little modification, most existing geometry-based protocol can be easily combined with the FHS scheme. For each broadcasting host, the selection of forwarding hosts can be done in $O(n)$ time, where n is the number of neighboring hosts of the host. Thus, the FHS scheme is more efficient in time complexity than most topology-based protocols [95][42] which are at least bounded by $O(n^2)$. Moreover, since the overheads for carrying the IDs of selected forwarding hosts are small, and fast forwarding hosts can increase broadcast speed, the FHS scheme is scalable for large ad hoc networks.

5.4 Performance Evaluation

To evaluate the performance of the FHS scheme, simulation experiments have been conducted in a homogeneous network. All hosts have the same communication range with radius $r = 1$ unit. Hosts are randomly distributed in an area of $S = 30 \times 30$ square units, and the number of hosts (N) in the network ranges from 1000 to 5000. For each given number of hosts, 10 random static graphs are generated, and 1,000 broadcast requests are simulated for each graph. Mobile hosts share a single common channel using a simplified version of IEEE 802.11 DCF [53] without RTS/CTS. To prevent the border effect [62] of the area with finite space, the broadcast requests are only launched by the hosts in the central area from which the broadcast packets cannot reach the network borders through the maximum TTL hops.

Table 5-1. Simulation Parameters for broadcast protocols

Items	Value
MaxDeferTime	0.1 sec
Defer Time Slot	0.01sec
D_{th} (Distance-based protocol)	0.6 unit
r_d (FHS scheme)	0.4 unit
TTL	5

A discrete simulator is developed to compare the performance of four broadcast relay protocols, i.e., the distance-based protocol, the angle-based protocol, the FHS distance-based protocol and the FHS angle-based protocol. Each protocol uses the distance-based defer time [50]. Table 5-1 shows the applied simulation parameters. Based on simulation experimentation, the threshold r_d in FHS is set as 0.4 unit for both FHS distance-based protocol and FHS angle-based protocol. The performance metrics to be observed are:

- Delivery ratio: the number of reachable hosts using a specific broadcast protocol over the number of the maximum reachable hosts using blind flooding mechanism with a certain TTL hops.
- The number of forwarding hosts: the average number of hosts which take part in broadcasting.

Figure 5-12 shows the delivery ratio versus the number of hosts in the network. It is clear that the FHS scheme is able to help the distance-based protocol to significantly increase the delivery ratio. For example, the delivery ratio for the FHS distance-based protocol is about 25% more than that for the distance-based protocol with $N = 1000$ hosts. This is because that the transmissions from the selected fast forwarding hosts h_{Fi} cover more areas, and the compensative forwarding hosts h_C are able to relay packet even in a sparse network. The

delivery ratio for both the FHS distance-based protocol and the distance-based protocol increase when the number of hosts in the network increases. On the other hand, both the angle-based protocol and the FHS angle-based protocol always achieve 100% reach-ability regardless of the host density.

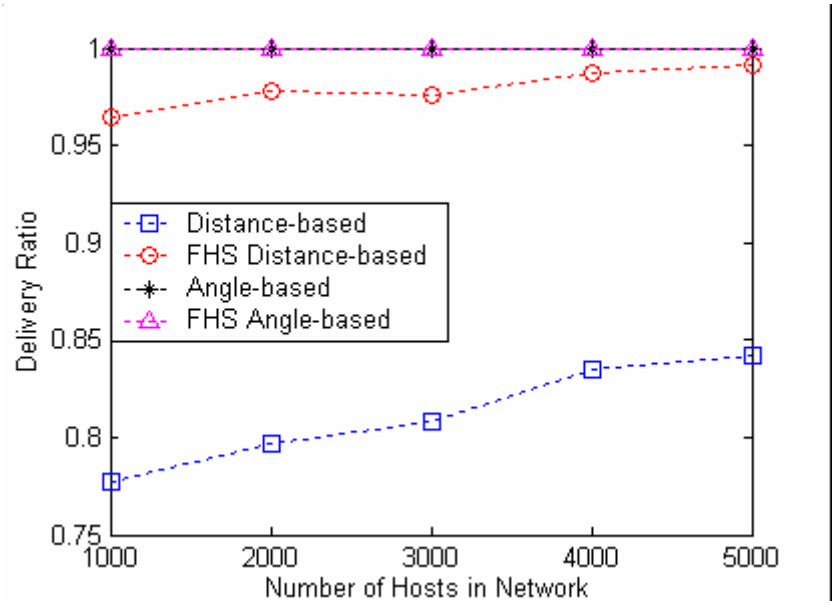


Figure 5-12. The average delivery ratio versus the number of hosts N

Figure 5-13 shows the average number of forwarding hosts required by each broadcast protocol. It illustrates that each protocol except blind flooding is scalable in terms of higher host density. It can be seen that the FHS scheme is able to enhance the angle-based protocol by reducing a large amount of forwarding hosts. For example, when the number of hosts in the network is 5000 and TTL=5, the number of forwarding hosts for the FHS scheme combined with the angle-based protocol is 30% less than that for the angle-based protocol. The effect of FHS scheme on reducing the redundant rebroadcasts increases when the network host density increases. This is because of more available forwarding hosts in the symmetrical area

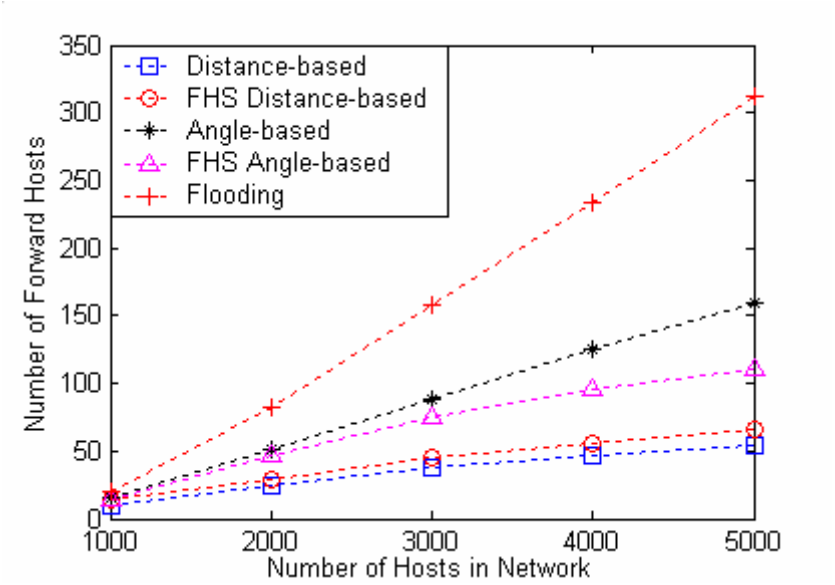


Figure 5-13. The average number of forwarding hosts versus the number of hosts N

On the other hand, the distance-based protocol and the FHS distance-based protocol generate the least number of forwarding hosts among the four geometry-based protocols. Referring to Figure 5-12 and Figure 5-13, it can be found although the number of forwarding hosts for the FHS distance-based protocol is slightly larger than that for the distance-based protocol, FHS does not use the increase of the number of rebroadcasts as the trade off to improve the delivery ratio for distance-based protocol.

A separate simulation for the dynamic environment is conducted to compare the stability of different broadcast protocols. Every host moves according to a uniformly distributed velocity, i.e., the moving rate and direction are uniformly distributed among predefined ranges $[0, V_m]$ and $[0, 2\pi]$ respectively, where V_m is the maximum moving rate. Moreover, each host sends a hello packet to its neighboring hosts to exchange location information once every 0.5 second. The simulation runs 100 times and generates an average value for each case. Considering that the Dominant Pruning (DP) broadcasting protocol has the highest

performance among the existing topology-based protocols [42], it is selected to compare the delivery ratio with the geometry-based protocols in the dynamic environment.

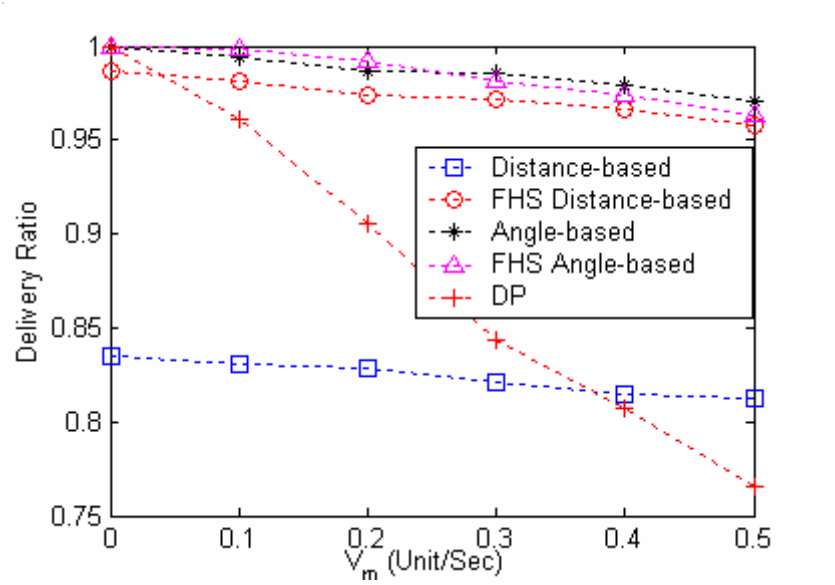


Figure 5-14. The effect of host mobility on delivery ratio

$N = 4000$

Figure 5-14 shows the delivery ratio versus the maximum host moving rate V_m when $N = 4000$. It can be found that when V_m has a fixed value, the FHS distance-based protocol always outperforms the distance-based protocol, and the FHS angle-based protocol and the angle-based protocol have the similar values of delivery ratio. From Figure 5-14, it also can be seen that the DP protocol are more sensitive to host mobility than the four geometry-based protocols. As the geometry-based protocols use the location information of direct neighboring hosts to make broadcasting decision, topology changes have little influence on these protocols. The delivery ratios for the geometry-based protocols drop a little while V_m increases, since the position information provided by hello packets may be stale in the network with high host mobility. By contrast, the delivery ratio for the DP protocol drastically decreases with the increase of V_m. The DP protocol has the lowest delivery ratio

among all the five protocols when $V_m \geq 0.4$ Unit/Sec. This is because under the DP protocol, the required location information in 2-hop neighborhood cannot be updated in time when local topology changes significantly. Therefore, comparing to topology-based broadcasting protocol, the geometry-based protocols combining with FHS has better reliability in highly dynamic ad hoc networks.

5.5 Conclusion

Firstly, this chapter presents the conditions to achieve the upper bound of broadcast coverage in an ad hoc network, which is useful for the design of suboptimal broadcast relay scheme to improve the performance of geometry-based protocols. Secondly, this chapter presents a novel FHS scheme, which is able to forward packets effectively, especially when network host density is heavy. Simulation shows that the broadcast performance improved by the FHS scheme increases when the network host density increases. The geometry-based protocols combined with FHS works reliable for highly dynamic ad hoc networks. Furthermore, the FHS scheme is able to help the distance-based protocol to increase the delivery ratio while keeping low number of forwarding hosts. On the other hand, FHS is able to enhance the performance of the angle-based protocol by reducing a large amount of redundant rebroadcasts while maintaining high delivery ratio. With only a little extra overhead and computational load, FHS can be easily combined with many existing geometry-based broadcast protocols, and improve the scalability of geometry-based broadcast protocols for large ad hoc networks.

Chapter 6 IEEE 802.11 MAC Protocol Enhanced by Busy Tones

6.1 Introduction

The unique characteristics of the MANETs make the design of an efficient MAC protocol very challenging. Since an ad hoc network lacks fixed infrastructure, distributed random MAC protocols, such as the Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA) [97], are usually used. However, the use of such randomly distributed MAC protocol may cause the hidden terminal problem and the exposed terminal problem [3][4]. As IEEE 802.11 is the factual industry standard for Wireless LAN in CSMA family, so far a majority of research on MANETs has used IEEE 802.11 series as the MAC protocol. Unfortunately, since IEEE 802.11 series is not originally designed for MANETs [98], the RTS/CTS (Request to Send/Clear to Send) handshake defined in IEEE802.11 MAC protocol is not able to prevent the hidden terminal problem properly in multihop networks. It should be noticed that such problem becomes even more serious when the interference range becomes more significant in an ad hoc network. This chapter focuses on schemes, which are able to improve the performance of IEEE 802.11 in the ad hoc network environment with large interference range.

6.1.1 CSMA/CA Protocols for Ad Hoc Networks

In a multihop ad hoc network, as shown in Figure 6-1, mobile host p located inside of the radio transmission range of the receiving mobile host j but outside of the radio transmission range of the transmitting mobile host i is defined as the hidden terminal [3][4]. In this case, since the hidden terminal p has no information about the on-going transmissions carried by mobile host i, its possible access to the channel may interrupt the delivery of data packet to the receiving hosts and degrade the network utilization. On the other hand, mobile host m located inside of the radio transmission range of the transmitting mobile host i but outside of radio transmission range of the receiving mobile host j is defined as exposed terminals [4]. In this case, the exposed terminals are prevented from accessing the channel to avoid interference. For example, as shown in Figure 6-1, the transmission from host m to host n has to be deferred when host i is transmitting packet to host j. Although there is no collision, the network communication resource is wasted unnecessarily.

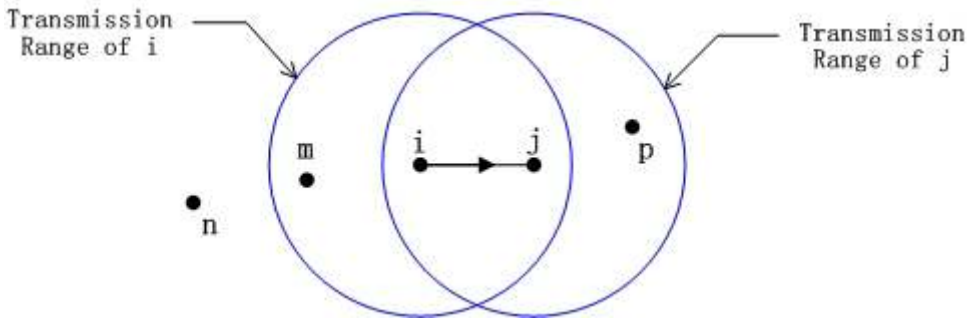


Figure 6-1. Hidden terminal and exposed terminal

In recent years, many MAC protocols have been proposed to solve the hidden terminal problem as well as the exposed terminal problem in ad hoc networks, including Busy Tone Multiple Access (BTMA) [3], Dual Busy Tone Multiple

Access (DBTMA) [99][100], Multiple Access Collision Avoidance (MACA) [101], MACA-BI (MACA by Invitation) [102], Floor Acquisition Multiple Access (FAMA) [103][104] and IEEE 802.11 Distributed Coordination Function (DCF) [53]. A good review on the existing ad hoc media access protocols is presented in [105].

However, among the above mentioned MAC protocols, only the DBTMA protocol is able to solve the problems for both hidden terminals and the exposed terminals using two out-of-band busy-tones to indicate the on-going transmissions, in which one busy tone indicates the transmitting busy and the other indicates the receiving busy. When transmitting busy-tone is received by neighboring hosts, then these neighboring hosts cannot transmit any packets and ignore any packet received. On the other hand, when neighboring hosts receive the receiving busy-tone, then they are prohibited from transmitting. Under the DBTMA protocol, a communication channel is split into two sub-channels, i.e., data channel and control channel. The control channel is to transmit RTS/CTS dialogue signaling. Moreover, in [106], power control is deployed to further increase channel utilization. A sender uses an appropriate power level to transmit packets so as to avoid interrupting on-going packet receptions. However, the analytical and simulation results shown in both [100] and [106] are evaluated under the condition that the interference range of each mobile host is assumed to be the same as the transmission range. This ideal situation is difficult to match the ad hoc network environment. It is because that the interference range in ad hoc network is usually considered to be larger than the

transmission range. Therefore, the network performance of DBTMA in real ad hoc network environment may have less significance.

6.1.2 Large Interference Range Problem

A technical survey has shown that more and more researchers have been focusing on the effects of large interference range on the performance in ad hoc networking environment [98][107][108][109][110][111], where the term of large interference range represents that the interference range of a host is larger than its transmission range. In this case, there are no any MAC protocols that are able to achieve the performance as expected in theory. This is because that the designs of these MAC protocols are only based on a simple assumption that the interference range of mobile hosts is the same as their transmission range in ad hoc networks.

The interference range in open space environment is usually calculated using the two-ray ground path-loss model [112]. When the distance x between the transmitter and the receiver is larger than the radius of the Freznel zone [112], the receiving signal power is inverse proportional to x^4 and the interference range of mobile host is $r_I = 1.78x$ [109]. Any other mobile host that is located inside of this interference range and transmits packet to its destination may interrupt the reception of packets. In [107], two solutions are proposed to reduce the interferences for IEEE 802.11 used in an ad hoc network. One is the Conservative CTS Reply (CCR) scheme, in which mobile host only replies CTS for RTS request when the received signal power of the RTS packet is larger than a certain threshold. The drawback of such CCR scheme is that the effective transmission range is reduced to 31.4%, which leads to lower network connectivity. The other approach is the Receiving Beam Forming (RBF)

scheme, in which directional antenna is deployed to avoid the interference from the directions with lower antenna gain. This scheme gives better performance comparing to CCR scheme. However, additional hardware is not easy to be implemented in the real ad hoc network environment.

It has also been demonstrated that the space reserved by IEEE 802.11 for a successful transmission is far from optimal and depends on the distance x [111]. Therefore, the revised Virtual Carrier Sensing (VCS) mechanism for wireless LAN scenarios is proposed for the overactive scenario ($x < 0.36r$), where r is the radius of every host's radio transmission range. However, there has no mechanism available for the under-active scenario ($x > 0.56r$) with more interference. There are also many other approaches that have been proposed to improve the IEEE 802.11 protocol to increase the interference range. For example, Adaptive IEEE 802.11 MAC (AMAC) protocol [108] enhances mobile hosts to periodically record the sensed signal strength and compute a threshold to adapt the CTS reply range. The simulation results have shown that the AMAC scheme outperforms the CCR scheme in terms of reducing interference. The Interference Aware (IA) MAC [113] protocol is another example, in which information of the received signal power and the interference levels are inserted into the MAC control packet. However, to the best of our knowledge, there have been none of MAC schemes able to solve the problems of large interference range properly.

Combining the dual busy tones with power control and some technologies used in the IAMAC protocol, two novel MAC schemes to improve the performance of IEEE 802.11 protocol for large interference range are presented in this chapter. One is called Fixed Power Dual Busy Tone (FPDBT) MAC protocol, which uses dual busy tones to identify the maximum possible range of interference. The dual busy tones are transmitted only when the

interference range is larger than the transmission range. The other novel scheme is called Variable Power Dual Busy Tone (VPDBT) MAC protocol, in which busy tones are transmitted with adjustable power level so that they can exactly indicate the instantaneous interference range. Differing from DBTMA [99], the proposed FPDBT and VPDBT MAC is based on 802.11 DCF, where data packets and control packets share one single physical channel.

6.2 IEEE 802.11 Enhanced By Dual Busy Tones

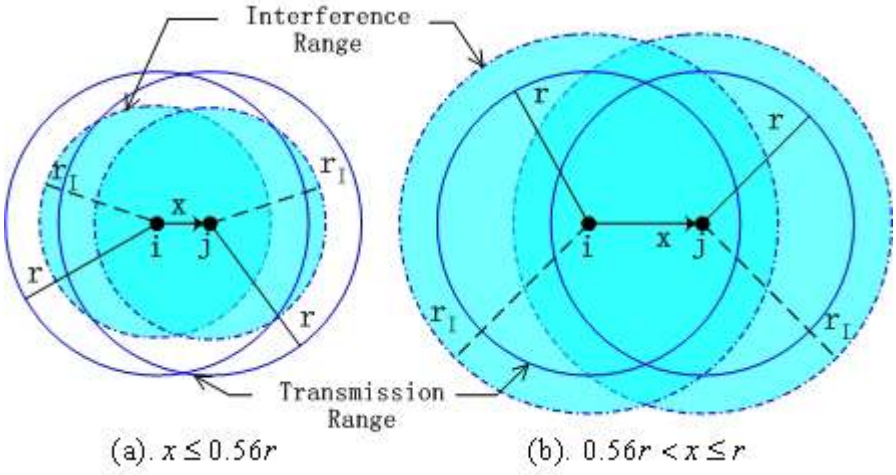


Figure 6-2. The interference ranges of host i and host j

In the ad hoc network environment, the RTS/CTS handshake of IEEE802.11 MAC protocol is not able to solve hidden terminal problem effectively, especially when the interference range is large. In fact, the interference range r_I for a mobile host in an ad hoc network is not fixed as the transmission range r of the mobile host. Instead, under two-ray ground path-loss model, the interference range depends on the transmitter–receiver distance x [107], which is given by $r_I = 10^{1/4} \cdot x \approx 1.78x$. As shown in Figure 6-2, in a two-dimensional homogeneous ad hoc network, a data packet is transmitted

from host i to host j, which have the same radio transmission range of radius r. When x is greater than $0.56r$, r_I is greater than r, and the hidden hosts located in the interference range may not be able to receive either RTS or CTS. Therefore, due to lacking of the duration information of the on-going transmission, these hidden hosts are not able to correctly set their Network Allocation Vector (NAV), which is used to indicate whether the channel is busy. In this case, if the hidden host transmits packet, the on-going transmission will be interrupted. Consequently, RTS/CTS handshake is not sufficient enough to reserve the interference area when the transmitter–receiver distance is larger than $0.56r$.

Two novel schemes, which are able to enhance the IEEE 802.11 MAC protocol to overcome the problems of hidden terminals and exposed terminals due to large interference range, are presented in this chapter. Figure 6-3 shows frequency allocation in the two proposed schemes. Two out-band busy tones, named as transmitting busy tone (BT_t) and receiving busy tone (BT_r), respectively, are placed at different frequency locations with sufficient frequency guard spaces in the signal channel, which is shared by data packets and control packets.

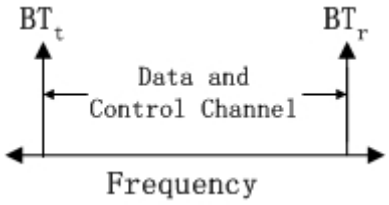


Figure 6-3. Frequency allocation

The following notations regarding power level are used in the presentation of the two novel schemes.

- P_t: the transmission signal power of packet at transmitter, which is assumed to be constant
- $P_r(x)$: the received signal power of packet at the distance x
- $P_{r0.56}$: the received signal power of packet at the distance $0.56r$
- P_{BTmin}: the minimum signal power level for mobile host to detect the busy tone.
- P_{BTmax}: the maximum signal transmission power of busy tone.
- P_{BT}: the transmission signal power of busy tone.

6.2.1 The Fixed Power Dual Busy Tone Scheme

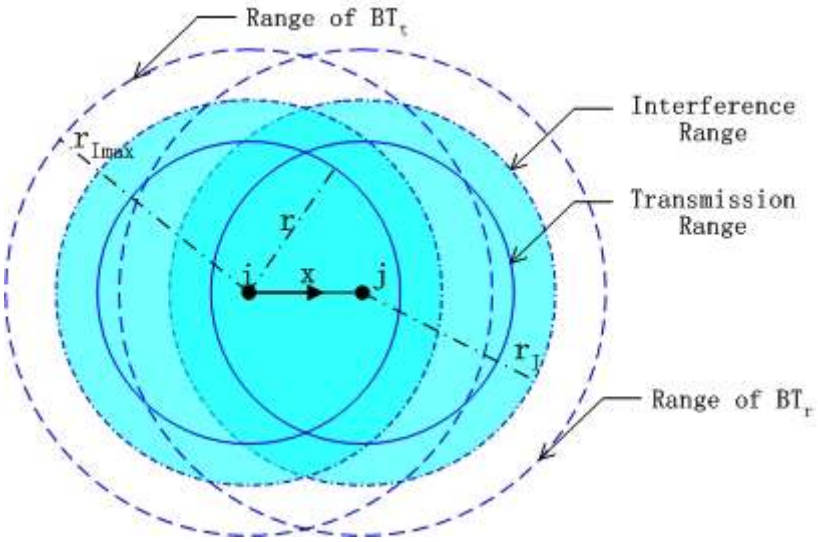


Figure 6-4. The coverage ranges of dual busy tones for FPDBT, $x > 0.56r$

Figure 6-4 shows the case that host i intends to transmit a data packet to host j at the distance x. When the distance x between i and j is larger than $0.56r$, the transmission range is smaller than the interference range, which is the shadow area shown in Figure 6-4. Since $0 \le x \le r$ and $r_I \approx 1.78x$, the maximum interference range is $r_{Imax} \approx 1.78r$, and the hidden hosts may exist inside of the range of radius between r

and r_{Imax}. Under the FPDBT scheme, mobile host is able to transmit packets only when the channel is idle as well as there is no busy tone sensed. In order to defer the possible transmissions of all hidden hosts duration the data transmission from host i to host j, the receiving busy tone BT_r should cover the maximum interference range of host j, and the transmitting busy tone BT_t should cover the maximum interference range of host i.

According to the two-ray ground path-loss model [112], the received signal power at the distance x is given by

$$P_r(x) = \frac{P_t G_t G_r h_t^2 h_r^2}{x^4} \tag{47}$$

where P_t is the transmission power of packet at transmitter, G_t and G_r are the antenna gains of transmitter and receiver respectively, h_t and h_r are the height of both antennas. When $x = 10^{-1/4} \cdot r \approx 0.56r$, the interference range equals to the transmission range. The received signal power at the distance $x \approx 0.56r$ is given by

$$P_{r0.56} = \frac{10 P_t G_t G_r h_t^2 h_r^2}{r^4} \tag{48}$$

As a busy tone must have enough power to cover the maximum interference range with radius r_{Imax}, if a busy tone is transmitted by the maximum power level P_{BTmax}, then the received power of busy tone at the distance r_{Imax} should at least equal the minimum power level P_{BTmin}, i.e.,

$$P_{BTmin} = \frac{P_{BTmax} G_t G_r h_t^2 h_r^2}{r_{Imax}^4}$$

Hence, $$P_{BTmax} = \frac{10 r^4 P_{BTmin}}{G_t G_r h_t^2 h_r^2} \tag{49}$$

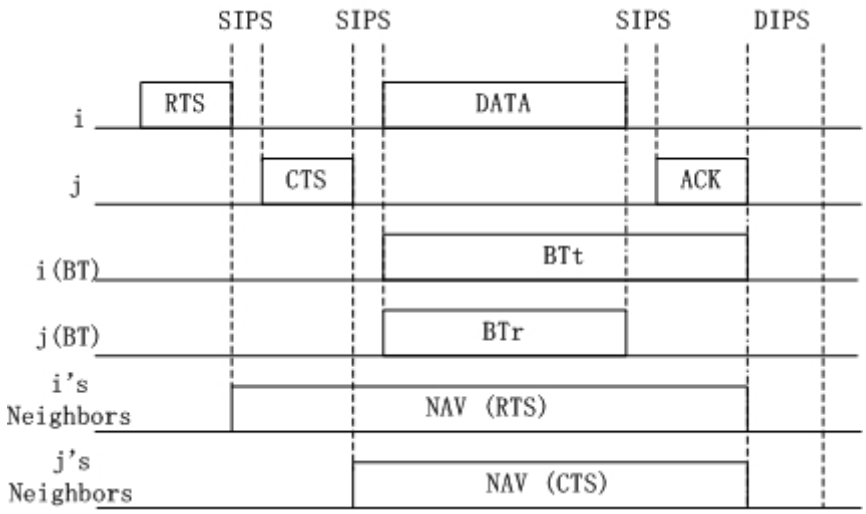


Figure 6-5. FPDBT transmission period, $x > 0.56r$.

The procedure of FPDBT can be described as follows: when host j (or i) receives a RTS (or RCT) packet from host i (or j), the host j (or i) checks if the receiving signal power $P_r(x)$ is smaller than the threshold $P_{r0.56}$. If $P_r(x) < P_{r0.56}$, i.e., the interference range is larger than the transmission range, then hosts i and j will transmit busy tones BT_t and BT_r, respectively. As shown in Figure 6-5, host j transmits BT_r by the maximum transmission power P_{BTmax} after it sends CTS to host i, and host i transmits BT_t by power P_{BTmax} during the transmission of the data packet and ACK (acknowledgment) packet. On the other hand, upon receiving a RTS or CTS packet, the neighboring hosts of hosts i and j set their NAV according to the duration information attached in the RTS/CTS packet. These neighboring hosts refrain their transmissions if BT_t or BT_r is detected, or their NAV indicates the current channel is busy. Therefore, when $x > 0.56r$, the FPDBT scheme uses the dual busy tones to identify the maximum interference range. In this case, both dual busy tones and VCS mechanism are applied to

combat interference. By contrast, if $P_r(x) > P_{r0.56}$, i.e., the interference range is less than the transmission range, host i and j simply deploy the IEEE 802.11 DCF.

Compared with IEEE 802.11 DCF, FPDBT MAC only needs additional dual busy tones transmitted by the maximum power P_{BTmax} when the interference range is larger than the transmission range. A FPDBT compliant host intending to transmit a packet must sense busy tone around it. Accordingly, those hosts, which cannot hear RTS or CTS, are able to sense busy tone and defer their transmissions during an on-going communication. For example, as shown in Figure 6-6, host m and n are located inside of the interference range of host i and the interference range of host j, but outside of the transmission range of host i and the transmission range of j, respectively. When the transmission from host i to j is in progress, the sole RTS/CTS handshake of IEEE 802.11 cannot prevent possible collisions caused by the transmission from host m (or n) to host p (or q). However, under the FPDBT scheme, the dual busy tones are able to prohibit any transmission by m (or n) during the on-going data transmission.

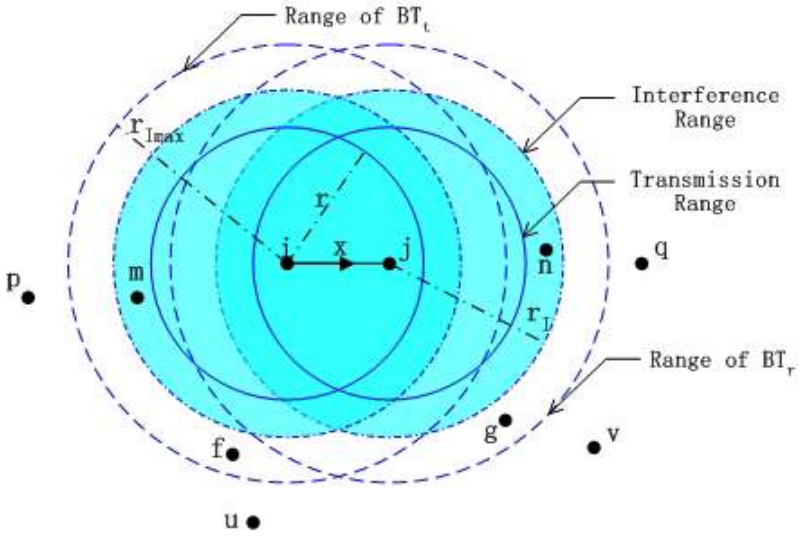


Figure 6-6. An example using FPDBT

It is obvious that FPDBT is able to solve the hidden terminal problem caused by large interference range. However, since the dual busy tones are always transmitted by the maximum power level, the spatial region reserved for the on-going traffic is larger than the interference range. For example, as shown in Figure 6-6, when $0.56r < x < r$, since hosts f and g are located outside of the interference ranges of hosts i and j, the transmission from host f (or g) to host u (or v) will not affect the communication between i and j. However, since f and g can detect the dual busy tones transmitted by i and j, any transmission from host f or g will be deferred unnecessarily. In this case, the exposed terminal problem becomes worse with respect to IEEE 802.11.

6.2.2 The Variable Power Dual Busy Tone Scheme

In order to improve the spatial reuse, another novel VPDBT scheme is proposed, where hosts are able to adjust the transmission power of busy tones so that busy tones can always indicate the interference ranges exactly. As shown in Figure 6-2, host i intends to transmit a data packet to host j at the distance x. Under VPDBT, during the transmission of a data packet, BT_r transmitted by host j should cover j's interference range, and BT_t transmitted by host i should cover i's interference range. Moreover, the VCS and physical channel sensing implemented in IEEE 802.11 DCF are totally replaced by busy tone sensing. A VPDBT compliant host can transmit a packet only if no busy tone is detected. Referring to Figure 6-7, the detail of VPDBT is formally described below.

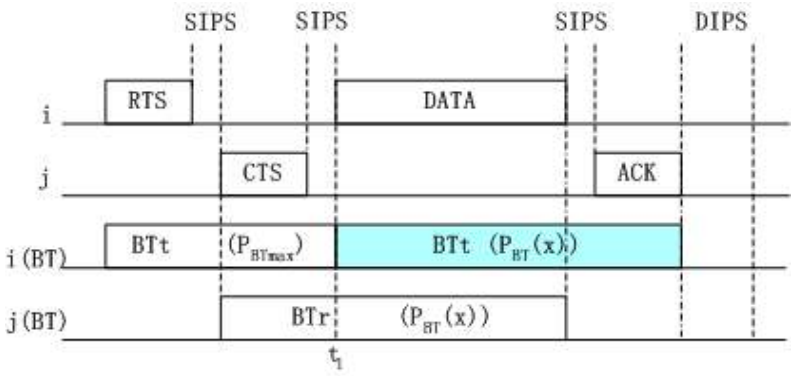


Figure 6-7. The transmission period of VPDBT

1) On a host i intending to send a RTS to host j, host i senses busy tones around it. If no busy tone is heard during a DIFS [7] period, host i sends RTS and launches the transmitting busy tone BT_t by the maximum power level P_{BTmax}.

2) Upon receiving i's RTS, host j senses receive busy tone BT_r around it. If BT_r is detected during a SIFS period, j ignores the RTS. Otherwise, j replies with a CTS, and transmits the receiving busy tone BT_r by transmission power level $P_{BT}(x)$.

3) After transmitting RTS, host i shall wait for a *CTSTimeout* interval. If a CTS can be correctly received duration this interval, host i starts to transmit the data packet, and adjusts the transmission power of BT_t to $P_{BT}(x)$. Otherwise, host i turns off BT_t and invokes a backoff procedure for retransmission after the *CTSTimeout* interval is expired.

4) After replying CTS, j prepares to receive the data packet. If the data packet can reach j in a SIFS period, j keeps on transmitting BT_r by the transmission power $P_{BT}(x)$

[7] In IEEE 802.11 MAC, Interframe Space (IFS) time intervals [53] are used to control priority access to the wireless medium. The Short IFS (SIFS) interval is the smallest IFS, followed by Point coordination function IFS (PIFS) and DCF-IFS(DIFS).

until the transmission of the data packet is completed. Otherwise, j can conclude that the RTS/CTS handshake is failure and turns off BT_r immediately.

5) Upon receiving the data packet, host j replies an ACK to confirm a successful transmission. On the other hand, after the data packet is transmitted, host i shall wait for an *ACKTimeout* interval. If an ACK packet from host j can be received during *ACKTimeout*, host i turns off BT_t and finishes this successful transmission. Otherwise, i turns off BT_t and invokes a backoff procedure for retransmission upon the expiration of *ACKTimeout*.

In the first step of VPDBT, when host i starts to transmit a RTS packet, it launches BT_t by the maximum power level P_{BTmax}, which can be determined by equation (49). The purpose of transmitting BT_t together with RTS is to inform all hosts inside of the maximum interference range of host i that a new RTS/CTS dialogue is on-going. Accordingly, any other transmission that may corrupt the reception of CTS at host i is prohibited by BT_t. On the other hand, during the transmission of data/ACK packet, the coverage range of BT_t must equals the interference range of hosts i, and the coverage range of BT_r must equals the interference range of host j exactly. In this case, let $P_{BT}(x)$ denote the transmission signal power of BT_t or BT_r, then the received signal power of BT_t or BT_r at the distance r_I should equals the minimum power level P_{BTmin}, i.e.,

$$P_{BTmin} = \frac{P_{BT}(x)G_tG_rh_t^2h_r^2}{r_I^4} = \frac{P_{BT}(x)G_tG_rh_t^2h_r^2}{10x^4} \tag{50}$$

where $r_I = 10^{1/4} \cdot x \approx 1.78x$.

Dividing (50) by (47), the transmission power of BT_t or BT_r is given by

$$P_{BT}(x) = \frac{10 P_{BTmin} P_t}{P_r(x)} \tag{51}$$

where P_t is the transmission power of packet at transmitter, and $P_r(x)$ is the received signal power of a RTS or CTS packet. Therefore, upon receiving a RTS (or CTS) packet, host j (or i) can easily determine the transmission power of BT_r (or BT_t) using equation (51).

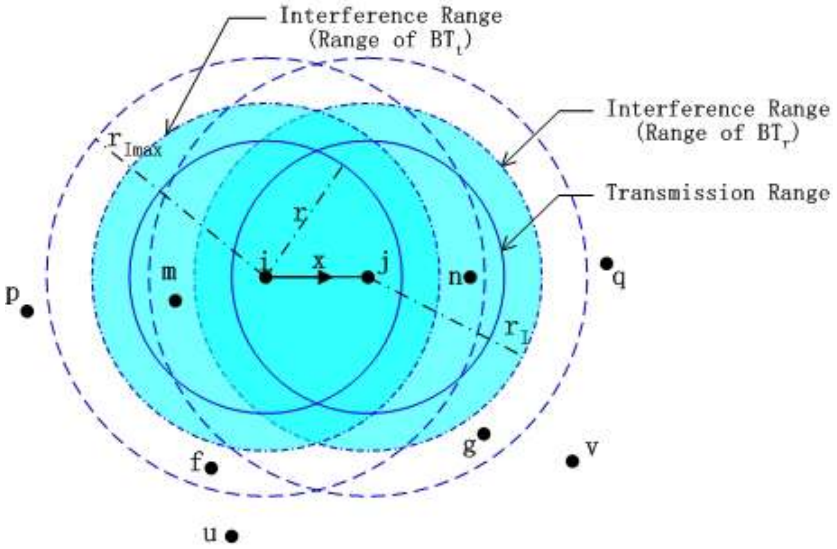


Figure 6-8. An example using VPDBT

Under VPDBT, since all hidden hosts located inside of the coverage ranges of dual busy tones will defer their transmission during an on-going communication, the hidden terminal problem can be easily solved by using VPDBT. On the other hand, as the coverage range of dual busy tone always equals the interference range, the hosts which are located outside of the interference ranges of the current communication pair, are allowed to access the channel. For example, as shown in Figure 6-8, under VPDBT, host f (or g) is able to send a packet to host u (or v) during the on-going transmission from host i to host j. However, such transmission is not allowed by using FPDBT. Consequently, the spatial reuse of VPDBT is better than that of FPDBT.

Furthermore, once a failing RTS/CTS handshake is detected, the VPDBT compliant transmitter and receiver immediately turn off their busy tones and allow other transmissions in the neighborhood. On the contrary, under IEEE 802.11, because of the VCS mechanism, an unsuccessful RTS/CTS handshake will prohibit the transmissions of neighboring hosts until the time interval reserved by the RTS/CTS is expired. As shown in Figure 6-8, host m is located inside of the transmission range of host i, and host n is located inside of the transmission range of host j. Using IEEE 802.11, a corrupt RTS (or CTS) from host i (or j) will defer the transmission from host m (or n) to host p (or q) for a long time period (more than the transmission time of a data packet). Thus the channel resource is wasted during this time period. However, if this case happens under VPDBT, hosts i and j will immediately turn off their busy tones and release the channel. Therefore, VPDBT can efficiently cope with the exposed problem encountered by IEEE 802.11.

6.3 Performance Analysis

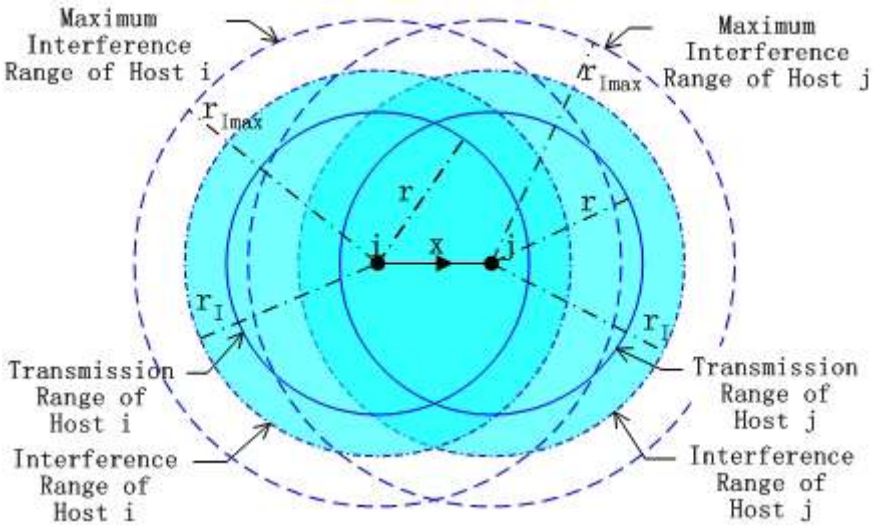


Figure 6-9. System model for performance analysis

In order to evaluate the performance of IEEE 802.11 DCF, FPDBT and VPDBT in an ad hoc network with large interference range, this section firstly focuses on the blocking area by each scheme. Secondly, the saturation throughput of each protocol is evaluated via a discrete Markov chain model. The following analysis considers the case that host i transmits a data packet to one of its neighboring hosts j at the distance x. Figure 6-9 shows the transmission ranges and the interference ranges of hosts i and j. The following notations regarding area are used in the performance analysis:

- S_i: the area covered by host i's radio transmission
- S_{Ii}: the area of host i's interference range
- $S_{i\cap j}$: the area overlapped by the transmission range of host i and the transmission range of host j
- S_{i-j}: the area covered by host i's transmission range but not covered by host j's transmission range
- S_{Ii-j}: the area inside of host i's interference range but not covered by host j's transmission range.
- S_{i-Ij}: the area covered by host i's transmission range but outside of host j's interference range
- S_{Ii-Ij}: the area inside of host i's interference range but outside of host j's interference range
- $S_{Ij-(i\cup j)}(x)$: the area inside of host j's interference range but not covered by the transmission range of either host i or host j
- S_{Imax}: the area of the maximum interference range of a host.

6.3.1 The Blocking Area

The blocking area is defined as the area which is reserved for the on-going traffic. For a given duration, no host located inside of the blocking area is allowed to transmit packets. Since the blocking area directly affect the maximum number of simultaneous transmissions in an ad hoc network, it can serve as an important benchmark to evaluate the spatial reuse of each scheme. From the fact that the duration of data transmission is much longer than that of RTS/CTS handshake, the area reserved for data transmission can simply represent the blocking area.

Let $S_{B\_802.11}$, S_{B_FPDBT} and S_{B_VPDBT} denote the blocking area of IEEE 802.11 DCF, FPDBT and VPDBT MAC, respectively, which are sensitive to the transmitter-receiver distance x. Referring to Figure 6-9, since the space reserved by IEEE 802.11 is the area which the RTS/CTS handshake covers, the blocking area of IEEE 802.11 DCF equals the transmissions coverage ranges of hosts i and j, that is

$$S_{B\_802.11}(x) = S_i + S_{i-j}(x) = 2\pi r^2 - S_{i\cap j}(x) \tag{52}$$

According to [114], $S_{i\cap j}(x) = 2r^2 q(\frac{x}{2r})$,

where $q(t) = \arccos(t) - t\sqrt{1-t^2}$.

Since host i chooses any neighboring host j as its receiver with equal probability, which is uniformly distributed in j's transmission range, the PDF of distance x between host i and the receiver j is given by

$$f_X(x) = \frac{2\pi x}{\pi r^2} = \frac{2x}{r^2} \tag{53}$$

Consequently, the average blocking area of IEEE 802.11 DCF can be given by

$$E(S_{B\_802.11}) = \int_0^r f_X(x) S_{B\_802.11}(x)dx = \frac{2}{r^2}\int_0^r x(2\pi r^2 - 2r^2 q(\frac{x}{2r}))dx$$

$$\approx 1.41\pi r^2 \tag{54}$$

On the other hand, the coverage ranges of dual busy tones determine the blocking area of VPDBT, which always equals the interference range, that is:

$$S_{B\_\mathrm{VPDBT}}(x) = S_{Ii} + S_{Ii-Ij}(x) = 2\pi r_I^2 - S_{Ii\cap Ij}(x) = 2\pi r_I^2 - 2r_I^2 q(\frac{x}{2r_I})$$

$$= 2\sqrt{10}x^2\left[\pi - q\left(\frac{1}{2}\cdot 10^{-1/4}\right)\right] \approx 13.44x^2 \tag{55}$$

The average blocking area of VPDBT is obtained as

$$E(S_{B\_\mathrm{VPDBT}}) = \int_0^r f_X(x) S_{B\_\mathrm{VPDBT}}(x)dx \approx 2.14\pi r^2 \tag{56}$$

Under the FPDBT scheme, if $x \le 0.56r$, the blocking area is determined by the RTS/CTS handshake; otherwise, the blocking area is determined by the coverage ranges of busy tones, which equal the maximum interference range of radius r_{Imax}. Using equation (52) and (55), the blocking area of FPDBT is given by

$$S_{B\_\mathrm{FPDBT}}(x) = \begin{cases} 2\pi r^2 - S_{i\cap j}(x), & x \le 0.56r \\ 2\pi r_{Imax}^2 - 2r_{Imax}^2 q(\frac{x}{2r_{Imax}}), & x > 0.56r \end{cases}$$

$$= \begin{cases} 2\pi r^2 - 2r^2 q(\frac{x}{2r}), & x \le 0.56r \\ 2\sqrt{10}r^2\left[\pi - q\left(\frac{x}{2\cdot\sqrt[4]{10}r}\right)\right], & x > 0.56r \end{cases}$$

The average blocking area of FPDBT is obtained as

$$E(S_{B\_\mathrm{FPDBT}}) = \int_0^r f_X(x) S_{B\_\mathrm{FPDBT}}(x)dx \approx 3.17\pi r^2 \tag{57}$$

According to equation (54), (56) and (57), it can be found that the average blocking area of IEEE 802.11 is the smallest, followed by VPDBT and FPDBT. However, the space reserved by IEEE 802.11 for a successful transmission is far from optimal. Figure 6-10 shows the blocking area versus the relative distance x/r. It can be seen that the blocking area for each scheme always increase with the relative distance. As the blocking area of VPDBT always coincides with the interference range, S_{B_VPDBT} is the optimal value to balance the trade off between spatial reuse and collision avoidance. Comparing the blocking area of IEEE 802.11 with that of VPDBT, it can be found that when $x \le 0.56r$, S_{B_VPDBT} is less than $S_{B\_802.11}$. In this case, the RTS/CTS dialogue in IEEE 802.11 may give false alarms to the hosts located outside of the interference range and reduce the spatial reuse. On the other hand, when $x > 0.56r$, S_{B_VPDBT} is larger than $S_{B\_802.11}$. However, the tight blocking area reserved by IEEE 802.11 cannot avoid collision efficiently (see the following subsection). From Figure 6-10, it also can be found that S_{B_FPDBT} is the largest block area among the three schemes. When $x \le 0.56r$, FPDBT has the same blocking area as IEEE 802.11. When $x > 0.56r$, S_{B_FPDBT} is much higher than the optimal value S_{B_VPDBT}, and the spatial reuse is largely reduced.

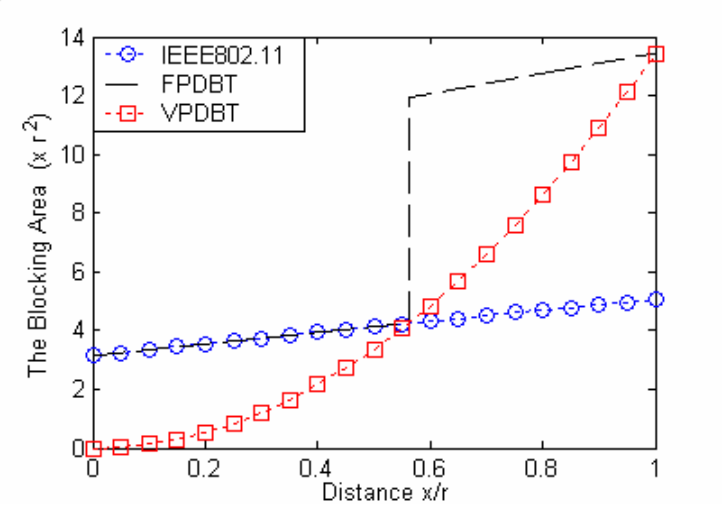


Figure 6-10. The blocking area versus the relative distance

6.3.2 The Saturation Throughput

The approximate model, which is developed by Takagi and Kleinrock [114] and further used by Wu and Varshney [115] in the study of CSMA and BTMA, is adopted here to evaluate the channel utilization for VPDBT and FPDBT. In [116] and [117], this model was also applied to estimate the performance of sender-initiated four-way handshake such as IEEE 802.11. In this chapter, this model is first extended to evaluate the saturation throughput of MAC protocols under the two-ray ground path-loss model.

It is assumed that mobile hosts are distributed in an ad hoc network as a two-dimensional Poisson point process [114][115] with density ρ. Let $\Pr(i,S)$ denote the probability that i hosts are found in a network with area S, then $\Pr(i,S)$ is given by:

$$\Pr(i,S)=\frac{(\rho S)^i}{i!}e^{-\rho S} \tag{58}$$

If all hosts have the same transmission range of radius r, the average number (the average degree) of neighboring hosts within a circular region of radius r can be obtained as $d=\rho\pi r^2$.

As the literatures [3][115] for CSMA and collision-avoidance protocols show, the performance of MAC protocols based on carrier sensing is much the same as the performance of their time-slotted counterparts in which the length of a time slot is much smaller than the transmission time of data packets. To make the analysis tractable, it is assumed that hosts operate in time-slotted mode in which the length of each time slot equals the SIFS interval defined in IEEE 802.11. The SIFS interval is denoted as τ, which includes

the propagation delay, the transmit-to-receive turn-around time, the carrier sensing delay and processing time. The transmission times of RTS, CTS, data, and ACK packets are normalized with regard to τ, and are denoted as t_{RTS}, t_{CTS}, t_{DATA} and t_{ACK}, respectively. Moreover, the throughput is derived based on the heavy traffic assumption, i.e., host i always has a packet in its buffer to be sent and the destination j is randomly chosen from one of its neighboring hosts. The probability p' that a host transmits in a time slot is independent at any time slot. Here p' is a protocol specific parameter, which depends on the channel's current state, the collision avoidance and backoff schemes being used. In the rest of this section, a discrete Markov chain is used to derive the approximate saturation throughput of IEEE 802.11 DCF, FPDBT and VPDBT, respectively.

A. IEEE 802.11 DCF

As shown in Figure 6-11, the transmission states of host i can be represented by a four-state Markov chain, where *Wait* is the state that host i defers for other hosts or backs off, *Succeed* is the state that host i can complete a successful four-way handshake, *Cfail* is the state that i initiates an unsuccessful RTS/CTS handshake, and *Dfail* is the state that the data/ACK dialogue is failure due to collision after a successful RTS/CTS handshake.

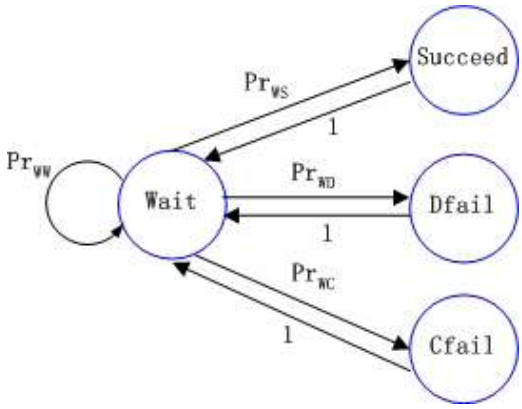


Figure 6-11. Markov chain model for IEEE802.11

In Figure 6-11, the event that host i continues to stay in *Wait* state happens when i does not initiate any transmission and none of *i*'s neighboring hosts initiates a transmission in one slot. The probability that a host does not transmit in a time slot is $1-p'$, and the probability that none of *i*'s neighboring hosts transmits in a slot is given by

$$\Pr_{NT} = \sum_{i=0}^{\infty}(1-p')^i \frac{d^i}{i!}e^{-d} = e^{-p'd}$$

Hence, the transition probability that i stays in wait state can be obtained as

$$\Pr_{WW} = (1-p')e^{-p'd}. \quad (59)$$

It is obvious that the duration of *Wait* state is $T_{Wait} = \tau$. Moreover, if no host is allowed to transmit data packets continuously, the transition probabilities from other states to *Wait* should be one.

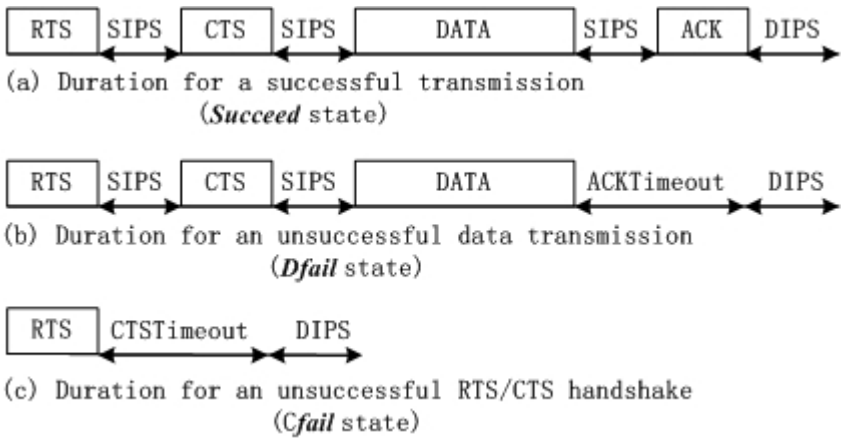


Figure 6-12. Time durations of the states using IEEE 802.11

Figure 6-12 shows the time durations of different states using IEEE 802.11, where $DIFS$, $ACKTimeout$ and $CTSTimeout$ [8] are the system parameters defined in IEEE 802.11 standard [53], and $SIFS = \tau$. According to Figure 6-12, the duration of each state is given by

$$T_{Succeed} = t_{RTS} + t_{CTS} + t_{DATA} + t_{ACK} + 3\tau + DIFS$$

$$T_{Dfail} = t_{RTS} + t_{CTS} + t_{DATA} + 2\tau + ACKTimeout + DIFS$$

$$T_{Cfail} = t_{RTS} + CTSTimeout + DIFS$$

Let $\Pr_{WS}(x)$ denote the probability that host i successfully completes a transmission to host j at the distance x, and $\Pr_{RTS}(x)$, $\Pr_{CTS}(x)$, $\Pr_{DATA}(x)$ and $\Pr_{ACK}(x)$ be the probabilities that the RTS, CTS, data and ACK packets are successfully received, respectively. Then $\Pr_{WS}(x)$ can be obtained as:

$$\Pr_{WS}(x) = p'(1-p') \cdot \Pr_{RTS}(x) \cdot \Pr_{CTS}(x) \cdot \Pr_{DATA}(x) \cdot \Pr_{ACK}(x), \quad (60)$$

where the term p' represents the probability that host i transmits in a slot, and the term $1-p'$ represents the probability that host j does not transmit in the time slot.

Let $\Pr_{WD}(x)$ denote the probability that the data/ACK dialogue is failure due to collision after a successful RTS/CTS handshake is completed, then $\Pr_{WD}(x)$ can be given by

$$\Pr_{WD}(x) = p'(1-p') \cdot \Pr_{RTS}(x) \cdot \Pr_{CTS}(x) \cdot \{1 - \Pr_{DATA}(x) + \Pr_{DATA}(x)[1 - \Pr_{ACK}(x)]\} \quad (61)$$

where the term $1 - \Pr_{DATA}(x)$ represents the probability that collision happens during the transmission of a data packet, and the term $\Pr_{DATA}(x)[1 - \Pr_{ACK}(x)]$ represents the probability that collision happens during the transmission of ACK.

---

[8] Under IEEE 802.11, after transmitting a RTS packet, the source host shall wait RTSTimeout amount of time without receiving a CTS packet before concluding that the RTS failed. Likewise, after transmitting a

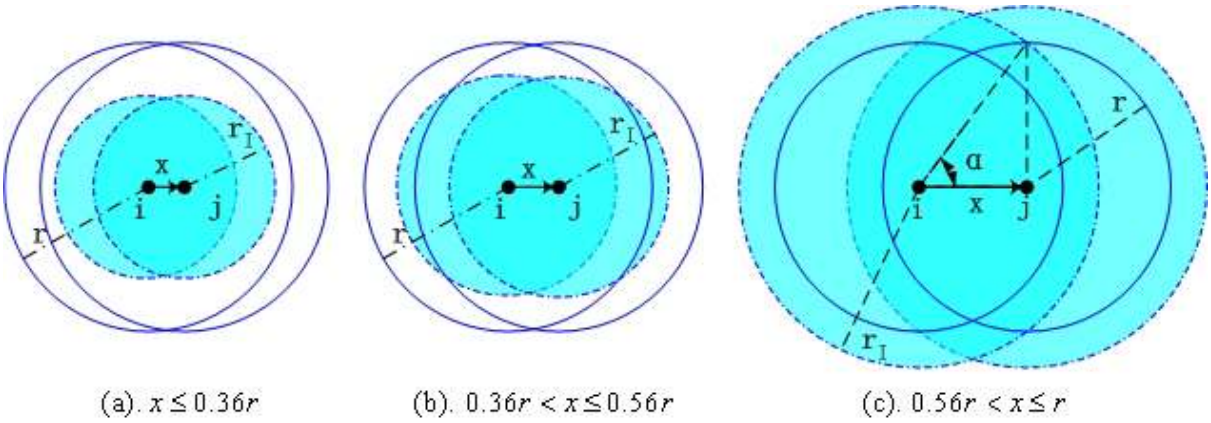


Figure 6-13. The transmission ranges and the interference ranges

However, the probabilities in equation (60) and (61) vary with the change of the distance x between hosts i and j, and they are dependent on both the transmission range and the interference range. As shown in Figure 6-13, the relationship between the transmission range (solid line circle) and the interference range (dotted line circle) can be categorized into three cases based on the range of the distance x. In case (a) (i.e., $x \le 0.36r$), the transmission range of host i or j can totally cover the interference ranges of the communication pair. In case (b) (i.e., $0.36r \le x \le 0.56r$), the transmission range is larger than the interference range, but the transmission range of host i or j cannot cover all the interference ranges of hosts i and j. Lastly in case (c) (i.e., $0.56r \le x \le r$),the interference range of host i or j totally contains its transmission range.

To make the analysis tractable, it is assumed that all collisions at hosts i and j are caused by the RTS packets transmitted from hidden hosts. The probability of further collisions is considered to be negligibly small. In order to successfully transmit a particular type of

---

data packet, the source host shall wait *ACKTimeout* amount of time without receiving an ACK packet before concluding that the data transmission failed.

packets between hosts i and j, i.e., RTS, CTS, data and ACK packet, the hosts located inside of a interfering area should not transmit packets for a certain time interval, named as vulnerable period. Table 6-1 lists the vulnerable periods and the interfering areas for transmitting the four different types of packets by using IEEE 802.11. The detail description can be found in Appendix E.2.

Table 6-1. The interfering area and vulnerable period for IEEE 802.11

Condition \ Packet Type		RTS	CTS	Data	ACK
$x \le 0.36r$	Vulnerable period	1 slot	--	--	--
	Interfering area	$S_{Ij}(x) = \pi r_I^2$	--	--	--
$0.36r < x \le 0.56r$	Vulnerable period	1 slot	--	--	--
	Interfering area	$S_{i\cap Ij}(x) = S_{Ij}(x) - S_{Ij-i}(x)$ $= \pi r_I^2 - \omega(r, r_I, x)$ \$	--	--	--
	Vulnerable period	$2t_{RTS} + 1$ slots	--	--	--
	Interfering area	$S_{Ij-i}(x) = \omega(r, r_I, x)$	--	--	--
$0.56r < x \le r$	Vulnerable period	1 slot	1 slot	$t_{RTS} + t_{DATA} + 1$ slots	$t_{RTS} + t_{ACK} + 1$ slots
	Interfering area	$S_{i\cap Ij}(x) = S_i(x) - S_{i-Ij}(x)$ $= \pi r^2 - \omega(r_I, r, x)$	$S_{(Ii-i)\cap j}(x)$ * $\approx \alpha(x)(r_I^2 - r^2)$	$S_{Ij-(i\cup j)}(x) \approx (r_I^2 - r^2)$ $\times\{\pi - \arccos[x/(2r)]\}$	$S_{Ii-(i\cup j)}(x) \approx (r_I^2 - r^2)$ $\times\{\pi - \arccos[x/(2r)]\}$
	Vulnerable period	$2t_{RTS} + 1$ slots	$t_{RTS} + t_{CTS} + 1$ slot	--	--
	Interfering area	$S_{Ij-i}(x) = \pi(r_I^2 - r^2)$ $+\omega(r_I, r, x)$	$S_{(Ii-i)-j}(x) \approx$ $[\pi - \alpha(x)](r_I^2 - r^2)$	--	--

Note: *Referring to Figure 6-13(c), $\alpha(x) = \arccos\dfrac{r_I^2 + x^2 - r^2}{2r_I x}$

\$ The expression of $\omega(r_A, r_B, x)$ can be found in the Appendix E.1.

Based on Table 6-1, the probabilities in equation (60) can be easily obtained. As shown in Figure 6-13, to successfully transmit a RTS packet from host i to j without any collision, there are three cases to be considered. When $x \le 0.36r$, none of hosts in the area $S_{Ij}(x) = \pi r_I^2$ should transmits for 1 slot, then $\Pr_{RTS}(x)$ can be obtained by

$$\Pr_{RTS}(x) = \sum_{n=0}^{\infty}(1-p')^n \frac{(\rho\pi r_I^2)^i}{i!} e^{-\rho\pi r_I^2} = e^{-p'\rho\pi r_I^2}, \quad x \le 0.36r .$$

When $0.36r < x \le 0.56r$, none of hosts in the area $S_{i\cap Ij}(x) = \pi r_I^2 - \omega(r, r_I, x)$ should transmit for 1 slot, and none of hosts in the area $S_{Ij-i}(x) = \omega(r, r_I, x)$ should transmit for $2t_{RTS} + 1$ slots. In this case, $\Pr_{RTS}(x)$ is given by

$$\Pr_{RTS}(x) = e^{-p'\rho[\pi r_I^2 - \omega(r,r_I,x)] - (2t_{RTS}+1)p'\rho\omega(r,r_I,x)}, \quad 0.36r < x \le 0.56r .$$

Likewise, when $0.56r < x \le r$, $\Pr_{RTS}(x)$ is given by

$$\Pr_{RTS}(x) = e^{-p'\rho[\pi r^2 - \omega(r_I,r,x)] - (2t_{RTS}+1)p'\rho[\pi(r_I^2 - r^2) + \omega(r_I,r,x)]}, 0.56r < x \le r .$$

Consequently, the probability that RTS can be successfully received can be obtained as

$$\Pr_{RTS}(x) = \begin{cases} e^{-p'\rho\pi r_I^2}, & x \le 0.36r \\ e^{-p'\rho[\pi r_I^2 - \omega(r,r_I,x)] - (2t_{RTS}+1)p'\rho\omega(r,r_I,x)}, & 0.36r < x \le 0.56r \\ e^{-p'\rho[\pi r^2 - \omega(r_I,r,x)] - (2t_{RTS}+1)p'\rho[\pi(r_I^2 - r^2) + \omega(r_I,r,x)]}, & 0.56r < x \le r \end{cases}$$

Likewise, the probabilities that CTS, data and ACK packets are successfully received can be given by

$$\Pr_{CTS}(x) = \begin{cases} 1, & x \le 0.56r \\ e^{-p'\rho\alpha(x)(r_I^2 - r^2) - (t_{RTS}+t_{CTS}+1)p'\rho[\pi - \alpha(x)](r_I^2 - r^2)}, & 0.56r < x \le r \end{cases}$$

$$\Pr_{DATA}(x) = \begin{cases} 1, & x \le 0.56r \\ e^{-(t_{RTS}+t_{DATA}+1)p'\rho(r_I^2 - r^2)\{\pi - \arccos[x/(2r)]\}}, & 0.56r < x \le r \end{cases}$$

$$\Pr_{ACK}(x) = \begin{cases} 1, & x \le 0.56r \\ e^{-(t_{RTS}+t_{ACK}+1)p'\rho(r_I^2 - r^2)\{\pi - \arccos[x/(2r)]\}}, & 0.56r < x \le r \end{cases}$$

Using equation (60) and (53), the transition probability $\Pr_{WS}$ can be expressed by

$$\begin{aligned} \Pr_{WS} &= \int_0^r f_X(x) \cdot \Pr_{WS}(x) dx \\ &= \frac{2p'(1-p')}{r^2} \int_0^r x \cdot \Pr_{RTS}(x) \cdot \Pr_{CTS}(x) \cdot \Pr_{DATA}(x) \cdot \Pr_{ACK}(x) dx \end{aligned} \tag{62}$$

Likewise, the transition probability $\Pr_{WD}$ from *Wait* state to *Dfail* state is derived by using equation (61), that is:

$$\Pr_{WD} = \frac{2p'(1-p')}{r^2}\int_0^r x \cdot \Pr_{RTS}(x) \cdot \Pr_{CTS}(x) \cdot \{1 - \Pr_{DATA}(x) + \Pr_{DATA}(x)[1 - \Pr_{ACK}(x)]\}dx, \quad (63)$$

Let π_W, π_S, π_C and π_D denote the steady-state probabilities of *Wait*, *Succeed*, *Cfail* and *Dfail* state, respectively. From Figure 6-11, it can be found that

$$\pi_W \Pr_{WW} + \pi_S + \pi_D + \pi_C = \pi_W$$

Using the fact that $\pi_S + \pi_D + \pi_C = 1 - \pi_W$, we have

$$\pi_W \Pr_{WW} + 1 - \pi_W = \pi_W$$

Using equation (59), then

$$\pi_W = \frac{1}{2 - \Pr_{WW}} = \frac{1}{2 - (1-p')e^{-p'd}}$$

Since $\pi_S = \pi_W \Pr_{WS}$, $\pi_D = \pi_W \Pr_{WD}$,

the steady-state probabilities of *Cfail* state is given by

$$\pi_C = 1 - \pi_W(1 + \Pr_{WS} + \Pr_{WD}).$$

Let $TH_{802.11}$ denote the throughput of the channel using IEEE 802.11. According to [115], the throughput equals the fraction of time in which the channel is engaged in the successful transmission of packets. Hence, $TH_{802.11}$ can be given by

$$TH_{802.11} = \frac{\pi_S t_{DATA}}{\pi_W T_{Wait} + \pi_S T_{Succeed} + \pi_D T_{Dfail} + \pi_C T_{Cfail}} \quad (64)$$

B. FPDBT MAC

As FPDBT MAC applies the same RTS/CTS handshake of IEEE802.11, the probabilities of successful reception of RTS and CTS packets using FPDBT are identical to those using

IEEE802.11. On the other hand, since that dual busy tones employed in FPDBT is able to avoid all collisions duration the transmission of data and ACK packets, the transmission states of host i using FPDBT can be represented by a three-state Markov chain, which consists of *Wait*, *Succeed* and *Cfail* state. The *Dfail* state shown in Figure 6-11 does not appears under FPDBT. In this case, the transition probability $\Pr_{WS}$ can be rewritten as

$$\Pr_{WS} = \frac{2p'(1-p')}{r^2}\int_0^r x \Pr_{RTS}(x)\cdot \Pr_{CTS}(x)dx \tag{65}$$

Since $\pi_S = \pi_W \Pr_{WS}$,

the steady-state probabilities of *Cfail* state is obtained as

$$\pi_C = 1-\pi_W-\pi_S = 1-\pi_W(1+\Pr_{WS})$$

Consequently, the throughput of the channel using FPDBT can be expressed by

$$TH_{\text{FPDBT}} = \frac{\pi_S t_{DATA}}{\pi_W T_{Wait} + \pi_S T_{Succeed} + \pi_C T_{Cfail}} \tag{66}$$

C. VPDBT MAC

Similar to FPDBT, the transmission states of host i using the VPDBT MAC scheme can be expressed by a three-state Markov chain, which consists of *Wait*, *Succeed* and *Cfail* state. However, under VPDBT, since the VCS and physical channel sensing are totally replaced by busy tone sensing, it is necessary to re-quantify the possible interference for an on-going transmission. As shown in Figure 6-7, the transmissions of CTS, data and ACK packets in VPDBT are always protected by dual busy tones. Hence, the possible collisions only happen when RTS is transmitted from host i to host j. The probability that host i successfully completes a data transmission to host j at distance x is given by

$$\Pr_{WS}(x) = p'(1-p')\cdot \Pr_{RTS}(x) \tag{67}$$

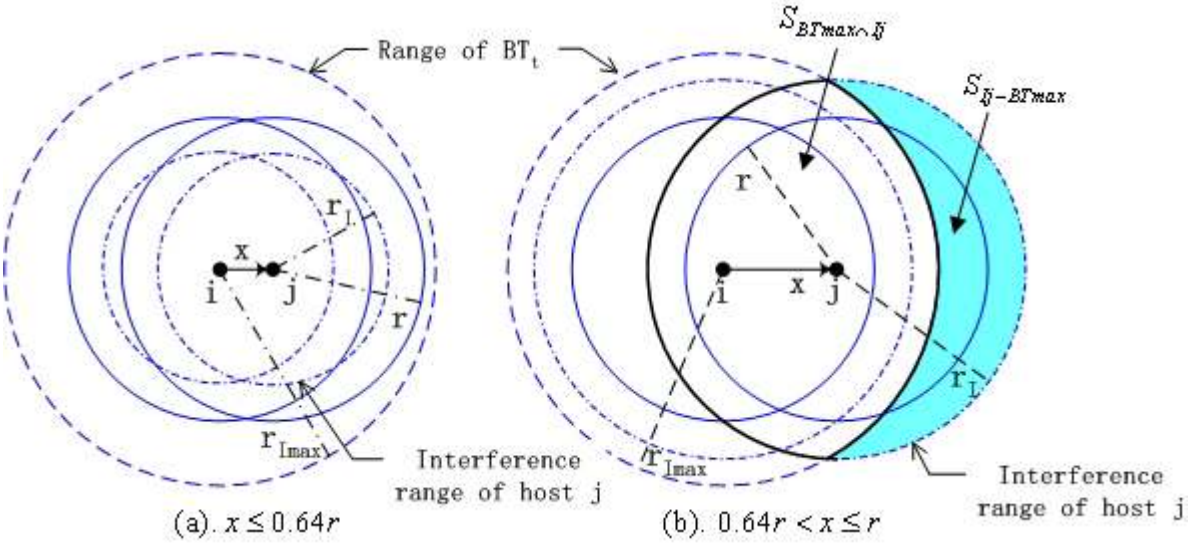


Figure 6-14. The transmission of RTS under VPDBT

When host i sends a RTS packet to host j, a transmitting bust tone BT_t is also launched by host i using the maximum power P_{BTmax}. In this case, the coverage range of BT_t is $r_{Imax} = 1.78r$. Figure 6-14 illustrates the following two cases: when $x \le 0.64r$, i.e., $x + r_I \le r_{Imax}$, BT_t is able to cover all the interference ranges of hosts i and j. To avoid the collision during the reception of RTS at host j, none of hosts in the interference area $S_{Ij} = \pi r_I^2$ should transmit for 1 slot. On the other hand, when $0.64r < x \le r$, i.e., $x + r_I > r_{Imax}$, BT_t can only cover a part of j's interference range, denoted as $S_{BTmax \cap Ij}$. As shown in Figure 6-14 (a), $S_{BTmax \cap Ij}$ is the area enveloped by the bold arcs, which is given by

$$S_{Imax \cap Ij} = \pi r_I^2 - \omega(r_{Imax}, r_I, x),$$

where the expression of $\omega(r_{Imax}, r_I, x)$ can be found in Appendix E.1. Obviously, hosts in the interference area $S_{BTmax \cap Ij}$ cannot transmit for 1 slot. However, those hosts located inside of j's interference range but outside of the coverage range of BT_t cannot hear either RTS or BT_t. In this case, only if hosts in the shadow area $S_{Ij-BTmax} = \omega(r_{Imax}, r_I, x)$ defer their

transmission for $2t_{RTS}+1$ slots, host j can receive the RTS from host i without any collision. Therefore, the probability for the successful reception of RTS can be expressed by

$$\Pr_{RTS}(x)=\begin{cases} e^{-p'\rho\pi r_I^2}, & x\le 0.64r \\ e^{-p'\rho[\pi r_I^2-\omega(r_{Imax},r_I,x)]-(2t_{RTS}+1)p'\rho\omega(r_{Imax},r_I,x)}, & 0.64r<x\le r \end{cases}$$

Using equation (67), the transition probability $\Pr_{WS}$ is given by

$$\Pr_{WS}=\int_0^r f_X(x)\cdot\Pr_{WS}(x)dx=\frac{2p'(1-p')}{r^2}\int_0^r x\Pr_{RTS}(x)dx \tag{68}$$

Moreover, the event that host i continues to stay in *Wait* state happens when i does not initiate any transmission and none of hosts in the area $S_{Imax}=\pi r_{Imax}^2$ initiates a transmission in one slot. Hence, the transition probability that i stays in wait state can be obtained as $\Pr_{WW}=(1-p')e^{-p'\rho\pi r_{Imax}^2}$.

Thus, the steady-state probabilities of *Wait* state is given by

$$\pi_W=\frac{1}{2-\Pr_{WW}}=\frac{1}{2-(1-p')e^{-\sqrt{10}p'\rho\pi r^2}}$$

Since $\pi_S=\pi_W\Pr_{WS}$,

the steady-state probabilities of *Cfail* state is obtained as

$$\pi_C=1-\pi_W-\pi_S=1-\pi_W(1+\Pr_{WS})$$

Consequently, the throughput of the channel using VPDBT can be expressed by

$$TH_{\text{VPDBT}}=\frac{\pi_S t_{DATA}}{\pi_W T_{Wait}+\pi_S T_{Succeed}+\pi_C T_{Cfail}} \tag{69}$$

The numerical results of the throughputs for the three MAC protocols are presented in the next section.

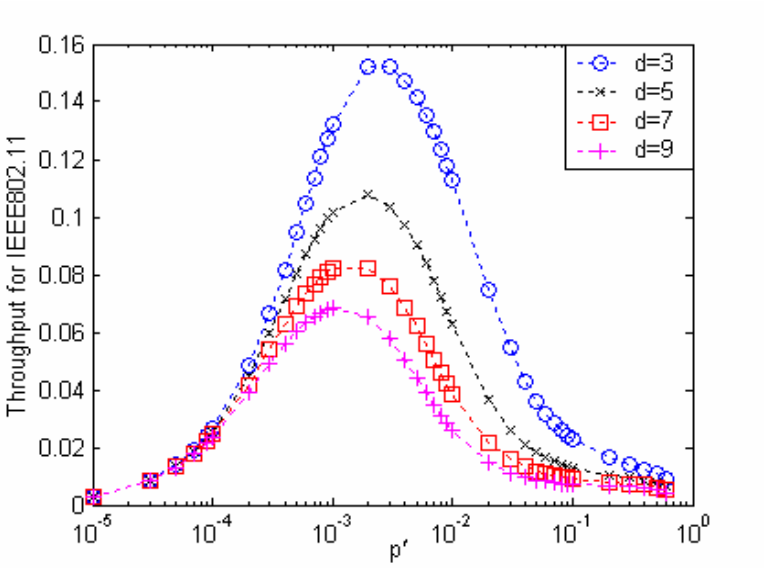
6.4 Numerical results and discussion

The numerical results of throughputs are evaluated by using Frequency Hopping Spread Spectrum (FHSS) parameters [118][119] for IEEE802.11 standard. Table 6-2 shows these parameters and their normalized values with regard to τ (SIFS).

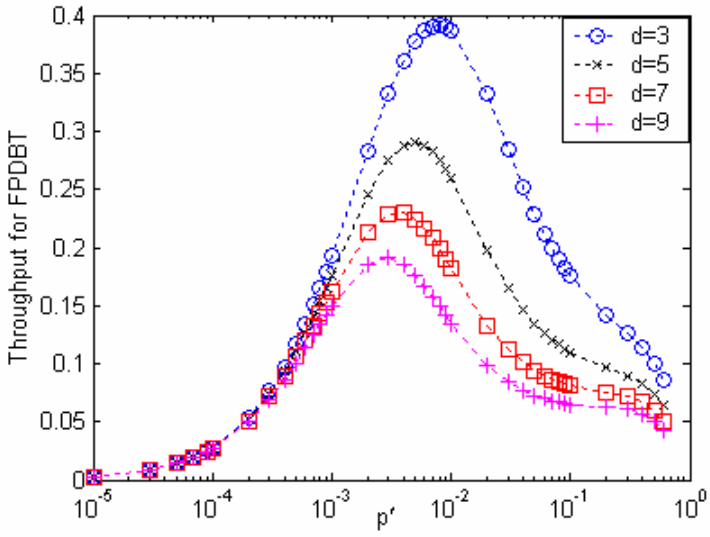
Table 6-2. FHSS system parameters

Duration	τ (SIFS)	DIFS	t_{RTS}	t_{CTS} , t_{ACK}	t_{DATA}	*CTSTimeout,ACKTimeout*
Actual Time	28 μ s	128 μ s	288 μ s	240 μ s	8184 μ s	300 μ s
Normalized	1	5	10	9	292	11

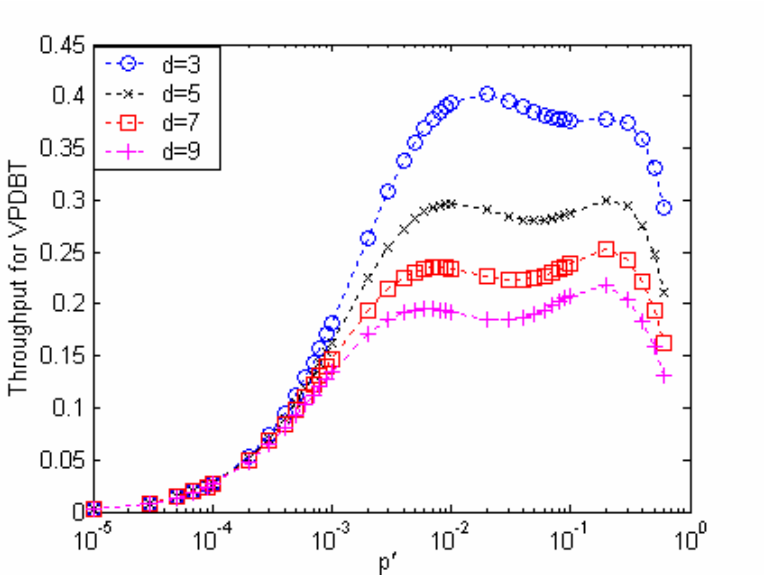
Figure 6-15 shows the throughputs versus the probability p' that a host transmits a packet in a time slot, by using IEEE 802.11, FPDBT and VPDBT. It can be seen that for each scheme, the throughput always decreases with the increase of d , when p' has a fixed value. This is because that the probability of collisions increases when the host density increases. Figure 6-15 also shows that FPDBT and VPDBT outperform IEEE 802.11 for various values of p'. For example, when $d = 3$, the maximum throughputs of FPDBT and VPDBT are about 2.6 times of that of IEEE 802.11. The reason is that FPDBT and VPDBT deploy dual busy tones to avoid possible collisions during the transmission of data and ACK packets. Contrarily, the RTS/CTS handshake used in IEEE 802.11 cannot completely solve hidden terminal problem. Comparing the throughput of FPDBT with that of VPDBT, it can be found that their performance is similar when the value of p' is small. However, when p' becomes large (more than 10^{-2}), the throughput for VPDBT still keeps large values, while the throughput of FPDBT becomes small. This is because the busy tones used in VPDBT is more effective, which can eliminate most collisions during the transmission of RTS and avoid all conflicts during the transmission of CTS. Therefore, under heavy traffic load, VPDBT shows better performance than the other two schemes.



(a). IEEE 802.11



(b). FPDBT



(c).VPDBT

Figure 6-15. Saturation Throughput

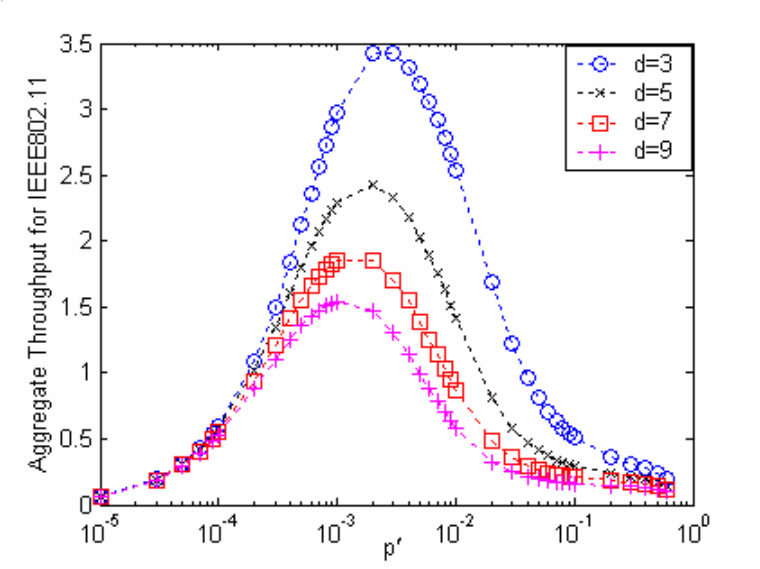
Furthermore, the aggregate throughput is used to evaluate the performance taking into account of both the channel throughput and the spatial reuse for each MAC scheme. The aggregate throughput is defined as the total throughput of simultaneous transmissions in the network, which can be approximated as

$$ATH \simeq TH \cdot \frac{S}{E(S_B)}, \tag{70}$$

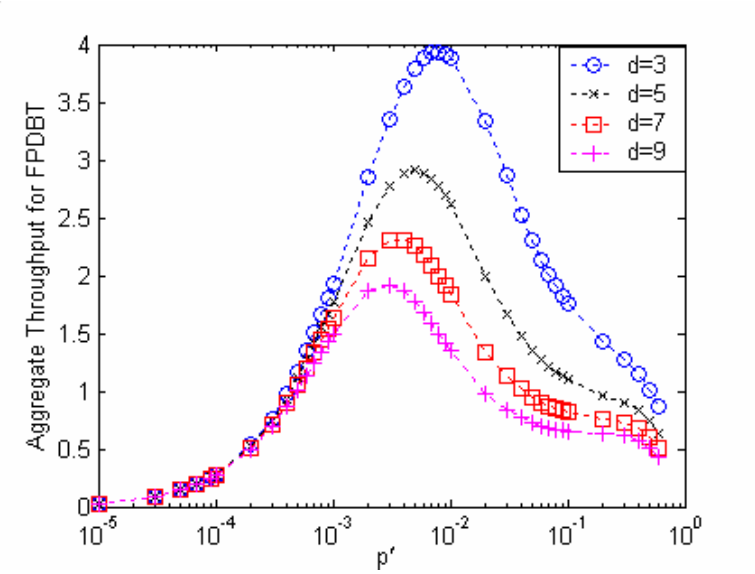
where TH is the saturation throughput of each transmitter-receiver pair, S is the area of the network, and $E(S_B)$ is the expected mean of the blocking area during each transmission. For IEEE 802.11, FPDBT and VPDBT schemes, $E(S_B)$ can be derived by equation (54), (56) and (57), respectively.

Figure 6-16 shows the aggregate throughputs versus p' for the three MAC schemes, when $r = 1$ unit and $S = 10 \times 10$ square units. It is clear when p' and d have fixed values, VPDBT has the largest aggregate throughput, followed by FPDBT and IEEE802.11. For

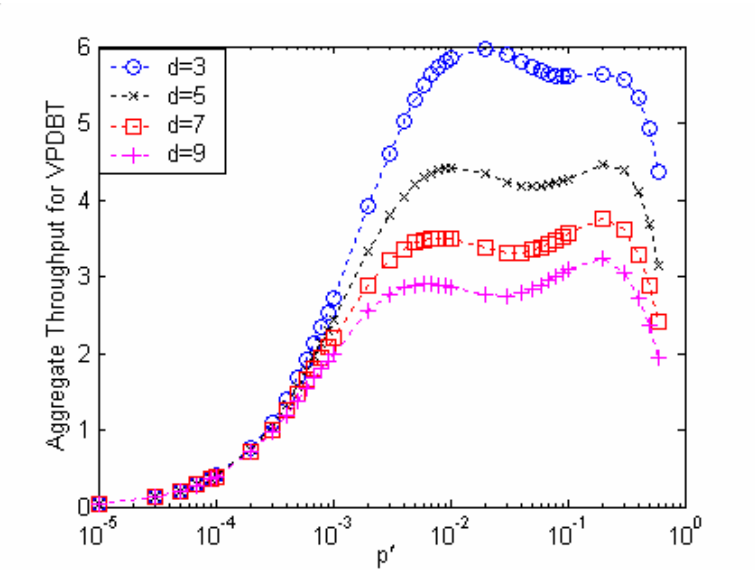
example, when $d = 3$, the maximum value of the aggregate throughput for VPDBT is 1.7 times of that for IEEE802.11, and 1.5 times of that for FPDBT. Compared with IEEE 802.11, VPDBT is able to increase the aggregate throughput 70%, while FPDBT can increase it 14%. However, from Figure 6-15, it can be found that both VPDBT and FPDBT can increase improve the throughput 160%. The improvement margin of the aggregate throughput is less than that of the throughput. This is because that the average blocking areas of VPDBT and FPDBT are larger than that of IEEE 802.11. Accordingly, the numbers of simultaneous transmissions using the two dual busy tone schemes are smaller than that using IEEE 802.11. However, since VPDBT can balance the trade off between spatial reuse and transmission collision, the performance of VPDBT can always achieve the best among the three schemes. It can be concluded that VPDBT is able to solve the hidden terminal problem and the exposed terminal problem in an ad hoc network with large interference range.



(a). IEEE 802.11



(b). FPDBT



(c).VPDBT

Figure 6-16. Aggregate throughput

$r = 1$ unit, $S = 10 \times 10$ square units

6.5 Conclusion

This chapter focuses on improving the performance of IEEE 802.11 in the open space environment with large interference range. Based on dual busy tones, two novel MAC schemes are proposed to reduce most interference under two-ray ground path-loss model. Among the two schemes, Fixed Power Dual Busy Tone MAC is a simple enhancement to IEEE 802.11 DCF. FPDBT uses dual busy tones with fixed transmission power to identify the maximum possible range of interference, when the interference range is larger than the transmission range. On the other hand, Variable Power Dual Busy Tone MAC use busy tone sensing to replace the virtual carrier sensing and physical channel sensing implemented in IEEE 802.11. The dual busy tones transmitted by adjustable power can exactly signify the interference range. Comparing to FPDBT, the implementation of VPDBT is more complicated, but VPDBT is able to achieve better network performance. Since the blocking area for VPDBT is less than that for FPDBT, the former has better spatial reuse. VPDBT can simultaneously solve the hidden terminal problem and the exposed problem which are encountered by IEEE 802.11.

Followed by the previous achievements, this chapter presents the first analysis model to derive the saturation throughput of CSMA MAC protocols under two-ray ground path-loss model in ad hoc networks. Moreover, we formulate the approximate throughputs and the blocking areas for IEEE 802.11, FPDBT and VPDBT. Numerical results show that VPDBT has the best performance, followed by FPDBT and IEEE 802.11. VPDBT and FPDBT can avoid all collisions during the transmissions of data packets and ACK. With respect to IEEE 802.11, both VPDBT and FPDBT improve the saturation throughput up to 1.6 times. Compared with IEEE 802.11, VPDBT is able to increase the

aggregate throughput 70%, while FPDBT can increase it 14% only. Both the VPDBT scheme and the FPDBT scheme can enhance the performance of IEEE 802.11 DCF in ad hoc networks with large interference range.

Chapter 7 Conclusion and Recommendation

7.1 Conclusion

This thesis focuses on the protocol design and performance analysis for mobile ad hoc networks. The first focus is on the network protocols to reduce the influence of host mobility. A novel location service protocol based on k-hop cluster is proposed to provide mobility management for large scale ad hoc networks. On the other hand, a novel broadcasting relay approach is proposed for effectively rebroadcast packets in an ad hoc network. The second focus is to improve the performance of IEEE 802.11 in the MAC layer for ad hoc networks.

Two mobility models including the handover model and the random walk model are used for the analysis of the characteristics of the host mobility in an ad hoc network. This is a very important step to fully understand the behavior of ad hoc networks using analytical methodology, since the host mobility is one of the most significant features in ad hoc networks. The performance analysis of mobility has been done for the handover model and the random walk model, respectively. Specially, the host random moving is studied in the case that the moving rate is assumed to be generally distributed and the direction is assumed to be uniform distribution. The numerical results obtained from the analysis have demonstrated that the random walk model is able to evaluate the characteristics of host mobility in an ad hoc network better than the handover model. However, the analysis of handover model is easier than the analysis of random walk model, especially, when mobile hosts move at higher speeds but less speed variation. The above analytical results can be

applied to the design of the MAC layer, the thresholds setting in routing protocols, the location information update in the location service protocols and the justification of simulation scenarios. Moreover, the numerical results of link changes can be used in the design of the suitable host density and the design of suitable transmission range of mobile hosts in an ad hoc network. Although the analysis in this chapter is only specific to certain mobility models, the methodology can also be applied to other models.

A novel k-hop Compound Metric Based Clustering (KCMBC) approach is proposed to keep cluster structure robust in the dynamic ad hoc network environment. KCMBC is a fast convergent and load balancing clustering approach. The time complexity of KCMBC is $O(k)$ rounds of messages. On the other hand, since KCMBC has taken into account of the host mobility, the clusters constructed by KCMBC are more stable than the other schemes. The results obtained from the simulation experiments show that the clusters created using the KCMBC approach have modest but more uniform cluster size. The cluster-head duration can be largely increased by KCMBC. Moreover, the redundant cluster architecture can tolerate false route and balances traffic load.

The design of robust and scalable location service protocol for large scale ad hoc networks is based on the proposed k-hop Clustering Based Location Service (KCBL) protocol, which utilizes the advantage of cluster architecture to construct the distributed location service system, where cluster-heads keep the location information of all clusters in the network and immediately reply the location enquiries from their cluster members. With the help of efficient inter-cluster location update mechanism, the KCBL protocol is able to provide more accurate location information in the destination's neighborhood and less accurate information for the hosts not in the destination's neighborhood. The frequency of inter-

cluster location updates can be determined by the cluster mobility pattern or the group characteristics. Therefore, the location updates is able to reflect the degree of cluster topology changes so that the control overheads can be significantly reduced. The numerical results obtained from simulation have indicated that KCBL is scalable in terms of various host density and host moving rate. The total cost of location management in the KCBL protocol is much less than the cost of the conventional link state protocol. It has been shown that the increase of k is able to increase the hit probability of location service and decrease the overhead in the initial stage as well as the total cost of location management. Moreover, the KCBL protocol is able to associate the network routing protocols to discovery the route path quickly with less overhead. The significant advantages of high ability to tolerate link breakage make the KCBL protocol are suitable for the design of scalable ad hoc networks.

The conditions to achieve the upper bound of broadcast coverage in an ad hoc network are obtained. It is able to provide useful information on the design of the suboptimum broadcast relay scheme to improve the performance of geometry-based protocols. A novel FHS scheme is presented. The FHS scheme is able to forward packets effectively, especially when network host density is heavy. The results obtained from simulations have shown that the FHS scheme is able to improve the broadcast performance significantly when the network host density increases. It is also demonstrated that the FHS scheme is able to enhance the distance-based protocol to increase the delivery ratio while the number of forwarding hosts is keeping low. On the other hand, FHS can improve the performance of the angle-based protocol by reducing a large amount of redundant rebroadcasts while maintaining high delivery ratio.

The IEEE 802.11 as the MAC layer with large interference range caused by the ad hoc network environment is studied. Two novel MAC schemes using dual busy tones are proposed to enhance the performance of IEEE 802.11 by reducing interference effects based on the two-ray ground path-loss model. The first proposed MAC scheme is called the fixed power dual busy tone MAC (FPDBT-MAC) scheme, in which the fixed power dual busy tones are used to indicate the maximum possible range of interference can be identified when the interference range is larger than the transmission range. The second proposed MAC scheme is called variable power dual busy tone MAC (VPDBT-MAC) scheme, in which the VCS and physical channel sensing used in IEEE 802.11 DCF are replaced by busy tone sensing, and the transmission power of busy tones can be adjusted to exactly cover the interference ranges of the transmitter-receiver pairs. The performance of the two proposed MAC schemes are evaluated, in terms of the approximate throughputs and the blocking areas. Numerical results show that both proposed VPDBT-MAC and FPDBT-MAC schemes are able to significantly enhance the IEEE 802.11 MAC layer protocol in an ad hoc network since they can avoid collisions for the transmissions of data packets and acknowledgement packets. On the other hand, both the VPDBT-MAC scheme and the FPDBT-MAC scheme are also able to improve the saturation throughput up to 160%. However, the VPDBT-MAC scheme is more effective than the FPDBT-MAC scheme to provide less blocking area and better spatial reuse. Therefore, the VPDBT-MAC scheme is able to simultaneously solve both of the hidden terminal problem and the exposed problem.

7.2 Recommendation and Future Research

The possible future studies based on the research works presented in this thesis are presented as follows:

Since the dynamic host mobility is one of the most significant features in ad hoc networks, the study of mobility has become an important topic in this area. Due to limited time, the modeling and analysis of host mobility in thesis has mainly been focused on the wireless link between two neighboring hosts. An interesting but also important future work is to extend the analytical model presented in Chapter 2 to the multiple hop situation on cluster structure, and study the effects of the host mobility on the performance of cluster architecture in terms of the sojourn time of host in a cluster, routing-path selection and network convergence time.

It is noticed that the proposed KCBL protocol associated with Mobile IP or IP can be used in support of inter-connection in heterogeneous wireless/wire networks, since the inter-connection between ad hoc networks and the other types of mobile/wireless networks is not a focus of this thesis. However, the inter-connectivity of ad hoc networks with the other types of networks is certainly an important issue to be studied in the future. On the other hand, security mechanism to protect private information within an ad hoc network needs to be addressed when the ad hoc network is inter-connected with the other networks. To offer effective host handover between different networks, the data encapsulation format among the different types of networks and also the route optimization techniques are open for future research.

Author's Publications Relevant to the Thesis

[1] Supeng Leng, Liren Zhang and *et al*, An Efficient Broadcast Relay Scheme for MANETs, To appear in Computer Communications, 2004

[2] Supeng Leng, Liren Zhang, An Enhanced IEEE 802.11 MAC Protocol for Mobile Ad Hoc Networks, Accepted by Computer Communications

[3] Supeng Leng, Liren Zhang and *et al*, The Location Service Protocol Based k-hop Clustering for Mobile ad hoc Networks, Submitted to IEEE Transactions on Vehicular Technology

[4] Supeng Leng, L. Zhang, Mobility Analysis for Random Walk model in ad hoc networks, Submitted to International Journal of Network Management

[5] Liren Zhang, Supeng Leng, k-hop Compound Metric Based Clustering Scheme for Mobile ad hoc Networks, Submitted to Computer Communications

[6] Supeng Leng, Liren Zhang and *et al*, IEEE 802.11 MAC Protocol Enhanced by Busy Tones, To appear in the Proceeding of IEEE ICC 2005, Korea

[7] Supeng Leng, Liren Zhang and *et al*, k-hop Compound Metric Based Clustering Scheme for Ad Hoc Networks, To appear in the Proceeding of IEEE ICC 2005, Korea

[8] Supeng Leng, Liren Zhang, Improving the Performance of IEEE 802 11 MAC Protocol for Multihop Wireless Networks, One of Best Paper Awards, Proceeding of IEEE Region 10 Infocom Colloquium on Broadband Access

Author's Other Publications

[1] Supeng Leng, K. R. Subramanian, N. Sundararajan, P. Saratchandran, Novel Neural Network Approach to Call Admission Control in High-speed Networks, International Journal of Neural Systems, Vol.13, No.4, pp251-262, 2003

[2] Supeng Leng, K. R. Subramanian, N. Sundararajan, P. Saratchandran, Call Admission control Using Minimal Resource Allocation Neural Network, Processing of CSNDSP' 2002, UK

References

[1] D. Hong, S. S. Rappaport, Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Nonprioritized Handover Procedures, IEEE Trans. Vehicular Tech., vol.35, no.3, pp.77-92, Aug. 1986

[2] T. Camp, J. Boleng, V. Davies, A Survey of Mobility Models for Ad hoc Network Research, Wireless Communication & Mobile Computing (WCMC), vol. 2, no. 5, pp. 483-502, 2002

[3] L. Kleinrock and F. Tobagi, Packet switching in radio channels, part II-the hidden terminal problem in carrier sense multiple access and the busy tone solution, IEEE Trans. on Commun., Vol.23, No.12, pp.1417-1433, 1975.

[4] P. Karn, MACA-A New Channel Access Method for Packet Radio. In Proc. 9th ARRL Computer networking Conference, pp. 134-140, 1990.

[5] B. M. Leiner, D. L. Nielson and F. A. Tobagi, Issues in Packet Radio Network Design. Proceedings of IEEE, Vol.75, No.1, pp.6-20, 1987

[6] J. Jubin and J.D. Tornow, The DARPA Packet Radio Network Protocols. Proceedings of the IEEE, Vol.75, No.1, pp.21-32, 1987

[7] A. Ephremides, J. Wieselthier, and D. Baker, A Design Concept for Reliable Mobile Radio Networks With Frequency Hopping Signaling, Proceedings of the IEEE, Vol.75, No.1, pp.56-73, 1987

[8] C. Shen and et al, Sensor Information Networking Architecture and Applications, IEEE Pers. Commun, Vol.8, No.4, pp. 52-59, Aug. 2001

[9] W. B. Heinzelman and et al, An Application-Specific Protocol Architecture for Wireless Microsensor Networks, IEEE Trans. on Wireless Commun., Vol.1, N0.4, pp.660-670, 2002

[10] I.F.Akyildiz and et al, A survey on sensor networks, IEEE Communications Magazine, Vol.40, No.8, pp.102-114, Aug. 2002

[11] Bluetooth Special Interest Group, Specification of the Bluetooth System 1.0b, Volume 1: Core, http://www.bluetooth.com, Dec. 1999

[12] J. Bray, C. F. Sturman, Bluetooth Connect Without Cables, Prentice Hall, 2001

[13] J. Misic, V. B. Misic, Bridges of Bluetooth County: Topologies, Scheduling and Performance, IEEE JSAC, vol.21, no.2, pp.240-258, 2003

[14] B. Crow et al., IEEE 802.11 Wireless Local Area Networks, IEEE Communications Magazine, vol. 35, no. 9, pp. 116-126, Sept. 1997

[15] ETSI, High Performance Radio Local Area Network (HIPERLAN), Draft Standard ETS 300 652, Mar. 1996

[16] H. Deng, W. Li, D. P. Agrawal, Routing Security in Wireless ad hoc Networks. IEEE Commun. Magazine, Vol.40, No.10, pp.70-75, Oct. 2002

[17] S. Capkun, L. Buttyan, J.-P. Hubaux. Self-organized public-key management for mobile ad hoc networks, IEEE Trans. on Mobile Computing, Vol.2, No.1, pp.52-64, 2003

[18] A. J. Goldsmith, S. B. Wicker, Design challenges for energy-constrained ad hoc wireless networks, IEEE Wireless Commun., vol.9, no.4, pp.8-27, 2002

[19] B.N.Clark, C.J.Colbourn and D.S.Johnson, Unit Disk Graphs, Discrete Math., Vol. 86, pp. 165-177, 1990

[20] A. Nasipuri, R. Castaneda, S. R. DAS, Performance of multipath routing for on-demand protocols in mobile ad hoc networks, Mobile Networks and Application, No.6, pp.339-349, 2001

[21] P. Santi, D.M. Blough, F. Vainstein, A probabilistic analysis for the radio range assignment problem in ad hoc networks, ACM MobiHoc 2001, pp.212-220, 2001

[22] C. Bettstetter and et al, Stochastic Properties of the Random Waypoint Mobility Model: Epoch length, direction distribution, and cell change rate, the proceeding of ACM MSWIM'02, pp.7-14, 2002

[23] J. Broch, D. Maltz, D. Johnson, Y. Hu, J. Jetcheva. Multihop wireless ad hoc network routing protocols. Proc. of the ACM/IEEE MOBICOM, pp.85–97, 1998

[24] J.J. Garcia-Luna-Aceves, E.L. Madrga. A multicast routing protocol for ad-hoc networks. In Proc. of INFOCOM, pp.784–792, 1999

[25] J.J. Garcia-Luna-Aceves, M. Spohn. Source-tree routing in wireless networks. In Proc. of the 7th International Conference on Network Protocols (ICNP), pp273, 1999

[26] P. Johansson et al, Routing protocols for mobile ad-hoc networks- a comparative performance analysis. In Proc. MOBICOM, pp.195–206, 1999

[27] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, J. Jetcheva, A performance comparison of multihop wireless ad hoc network routing protocols, Proc.ACM MobiCom'98, pp. 85-97, 1998.

[28] S. Basagni, I. Chlamtac, V.R. Syrotiuk, B.A. Woodward, A distance routing effect algorithm for mobility (DREAM), Proc. MOBICOM, pp.76-84,1998

[29] B. Karp, H. T. Kung, GPSR: Greedy perimeter stateless routing for wireless networks. Proc. of ACM MobiCom 2000, pp. 243-254, 2000

[30] Y. Ko., N.H. Vaidya, Location-aided routing (LAR) in mobile ad hoc networks, Wireless Networks, Vol.6, No.4, pp.307-321, 2000

[31] W. Liao, Y. Tseng, J. Sheu, Grid: A fully location-aware routing protocol for mobile ad hoc networks, Telecommunication Systems, Vol.18, No.1, pp.37-60, 2001

[32] J. Li and et cetra, A Scalable Location Service for Geographic ad hoc Routing, Proceedings of the 6th international conference on Mobile computing and networking, ACM, pp. 120 -130, 2000

[33] S. Capkun, M. Hamdi, J.P. Hubaux, GPS-free positioning in mobile ad-hoc networks, Proc. Hawaii Int. Conf. on System Sciences, pp.1-10, Jan. 2001

[34] J. Hightower, G. Borriello, Location Systems for Ubiquitous Computing, Computer, Vol. 34, No. 8, pp. 57–66, Aug. 2001

[35] J.J. Caffery, G.L. Stuber, Overview of Radiolocation CDMA Cellular Systems, IEEE Communications Magazine, pp. 38-45, April 1998

[36] J. Sharony, An architecture for mobile radio networks with dynamically changing topology using virtual subnets, ACM Mobile Networks and Applications, Vol.1, No.1, pp. 75-86, 1996

[37] Z.J. Haas and B. Liang, Virtual backbone generation and maintenance in ad hoc network mobility management, Proc. of IEEE INFOCOM 2000, pp. 1293-1302, 2000

[38] Z.J. Haas, M.R. Pearlman, The Performance of Query Control Schemes for the Zone Routing Protocol, IEEE/ACM Trans. on Networking, Vol.9, No.4, pp 427-438, 2001

[39] M. Gerla and J.T.C. Tsai, Multicluster, mobile, multimedia radio network, Wireless Networks, Vol.1, No.3, pp.255-265, 1995

[40] C. Chiang and M. Gerla, Routing and Multicast in Multihop, Mobile Wireless Networks, Proc. IEEE ICUPC97, pp.546-551, Oct. 1997

[41] G. Pei et al, A Wireless Hierarchical Routing Protocol with Group Mobility, Proc. IEEE WCNC99, pp. 53-60, Sept. 1999

[42] Jie Wu, Wei Lou, On reducing broadcast redundancy in ad hoc wireless networks, IEEE Trans. on Mobile Computing, Vol 1, Issue 2, pp.111 –122, 2002.

[43] H. Lim, C. Kim, Flooding in wireless ad hoc networks, Computer Communications, Vol.24, No 3-4, pp.353-363, 2001

[44] D. B. Johnson and D. A. Maltz, Dynamic Source Routing in ad hoc Wireless Networks, Mobile Computing, edited by Tomasz Imielinski and Hank Korth, Chapter 5, Kluwer Academic, pp. 153-181, 1996

[45] C. E. Perkins and E. M. Royer, Ad hoc On-Demand Distance Vector Routing, Proc. of the 2nd IEEE Workshop on Mobile Computing Systems and Applica., pp.90-100, 1999

[46] G. Pei, M. Gerla, T.-W. Chen, Fisheye State Routing: A Routing Scheme for ad hoc Wireless Networks, Proc. ICC 2000, pp. 70-74, June 2000

[47] B. Bellur, R. G. Ogier, A Reliable, Efficient Topology Broadcast Protocol for Dynamic Networks, Proc. IEEE INFOCOM '99, pp. 178—186, Mar. 1999

[48] P. Jacquet et al., Optimized Link State Routing Protocol, draft-ietf-manetolsr-05.txt, Internet Draft, IETF MANET Working Group, Nov. 2000

[49] Y. Tseng, S. Ni, E. Shih., Adaptive Approaches to Relieving Broadcast Storms in a Wireless Multihop Mobile ad hoc network, IEEE Trans. on Computers, Vol.52, No. 5, pp. 545–557, 2003

[50] M. Sun, W. Feng, T. Lai, Location aided broadcast in wireless ad hoc networks, GLOBECOM 2001, Vol. 5, pp. 2842-2846, 2001

[51] M. Sun, T.H. Lai, Computing Optimal Local Cover Set for Broadcast in ad hoc Networks, Proc. IEEE ICC 2002, pp. 3291-3295, 2002

[52] M. Sun, T. Lai, Location aided broadcast in wireless ad hoc systems, Proc. IEEE WCNC 2002, pp. 597-602, 2002

[53] ANSI/IEEE Std 802.11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, 1999

[54] S. Basagni, I. Chlamtac, V.R. Syrotiuk, Geographic Messaging in Wireless ad hoc Networks, Proc. 49th IEEE VTC, Vol.3, pp. 1957–1961, 1999

[55] A. Amis, R. Prakash, T. Vuong, and D.T. Huynh. Max Min D-Cluster Formation in Wireless Ad hoc Networks, Proc. IEEE INFOCOM, pp. 32-41, Mar. 1999

[56] I.D. Aron, S. Gupta, Analytical Comparison of Local and End-to-End Error Recovery in Reactive Routing Protocols for Mobile ad hoc Networks, Proc. ACM Workshop Modeling, Analysis, and Simulation of Wireless and Mobile Systems (MSWiM), pp. 69-76, 2000

[57] M. Grossglauser, D. Tse, Mobility Increases the Capacity of ad hoc Wireless Networks, Proc. IEEE Infocom, pp.477-486, Apr. 2001

[58] P. Gupta, P.R. Kumar, The Capacity of Wireless Networks, IEEE Trans. Information Theory, vol.46, no.2, pp.388-404, 2000

[59] P. Santi, D.M. Blough, The Critical Transmitting Range for Connectivity in Sparse Wireless ad hoc Networks, IEEE Trans. Mobile Computing, Vol.2, No.1, pp.25-39, 2003

[60] I. Stojmenovic, Location updates for efficient routing in ad hoc networks, Handbook of wireless networks and mobile computing, John Wiley & Sons, pp.451 – 471, 2002

[61] C. Bettstetter, G. Resta, P. Santi. The node distribution of the random waypoint mobility model for wireless ad hoc networks. IEEE Trans. on Mobile Computing, Vol.2, No.3, pp.257-269, 2003

[62] C. Bettstetter, Mobility Modeling in Wireless Networks: Categorization, Smooth Movement and Border Effects, ACM Mobile Comp. and Comm. Review, vol.5, no.3, pp. 55-67, 2001

[63] M. Zonoozi, P. Dassanayake, User Mobility Modeling and Characterization of Mobility Patterns, IEEE JASC, vol.15, no.7, pp.1239-1252, Sep. 1997

[64] K. Yeung, S. Nanda, Channel Management in Microcell/Macrocell Cellular Radio Systems, IEEE Trans. Vehicular Tech., vol.45, no.4, pp.601-612, Sep. 1996

[65] S. Nanda, Teletraffic models for urban and suburban microcells: Cell sizes and handover rates, IEEE Trans. Veh. Technol., vol.42, pp.673–682, Nov. 1993

[66] A. B. McDonald, T. Znati, A Mobility Based Framework for Adaptive Clustering in Wireless Ad hoc Networks, IEEE JSAC, vol.17, no.8, pp.1466-1487, Aug. 1999

[67] J.J. Garcia, E.L. Madrga, A multicast routing protocol for ad-hoc networks. IEEE INFOCOM, pp. 784–792, 1999

[68] G. Holland and N. H. Vaidya, Analysis of TCP performance over mobile ad hoc networks, in Proc. ACM MobiCom'99, pp.219-230, 1999

[69] S. R. Das, C. E. Perkins, and E. M. Royer, Performance comparison of two on-demand routing protocols for ad hoc networks, IEEE Personal Communications, Vol.8, No.1, pp.16-28, 2001

[70] L. Kleinrock, J. Silvester, Optimum transmission radii for packet radio networks or why six is a magic number, Proc. IEEE National Telecommunications Conference, pp 4.3.1-4.3.5, 1978

[71] I. Stojmenovic, Position-based routing in ad hoc network, IEEE Communic. Magazine, pp.128-134, July 2002

[72] M. Mauve, J. Widmer, H. Hartenstein, A Survey on Position-Based Routing in Mobile ad hoc Networks, IEEE Network Magazine Vol.15, No.6, pp.30-39, 2001

[73] E. M. Royer, C. K. Toh, A Review of Current Routing Protocols for ad hoc Mobile Wireless Networks, IEEE Personal Commun., Vol. 6, No. 2, pp.46-55, 1999

[74] D. M. Lazoff, A. T. Sherman, Expected Wire Length Between Two Randomly Chosen Terminals, SIAM Review, Vol. 37, No. 2, pp.235, June 1996

[75] D. M. Lazoff, A. T. Sherman, An Exact Formula for the Expected Wire Length Between Two Randomly Chosen Terminals, U. Maryland Baltimore technical report CS-94-08, July 1994

[76] Swades D. and et al, A Resource Efficient RT-QoS Routing Protocol for Mobile ad hoc Networks, WPMC'02, pp. 257-261, 2002

[77] K.N. Amouris, S. Papavassiliou, M. Li, A position based multi-zone routing protocol for wide area mobile ad-hoc networks, Proc. 49th IEEE VTC, pp.1365-1369, 1999

[78] C.R. Lin and M. Gerla, Adaptive clustering for mobile wireless networks, IEEE JSAC, Vol.15, No.7, pp. 1265-1275, 1997

[79] C. C. Shen and et cetra, CLTC_a cluster-based topology control framework for ad hoc networks, IEEE Trans. on Moble Computing, vol.3, no.1, pp.18-32,2004

[80] B. An, S. Papavassiliou, A mobility-based clustering approach to support mobility management and multicast routing in mobile ad-hoc wireless networks, Int. J. Network Mgmt., Vol.11, pp.387 – 395, 2001

[81] G. Chen, F. G. Nocetti, J. S. Gonzalez, I. Stojmenovic, Connectivity Based k-hop Clustering in Wireless Networks, Proc. The 35th Hawaii ICSS, pp.2450-2459, 2002

[82] M. Joa-Ng and I.T. Lu, A peer-to-peer zone-based two-level link state routing for mobile ad hoc networks, IEEE JSAC, Vol.17, No.8, pp.1415-1425, 1999

[83] KH Wang, B. Li. Group Mobility and Partition Prediction on Wireless Ad-Hoc Networks. In Proc. of IEEE ICC Conference, pp.1017-1021, 2002

[84] Z.J. Haas and B. Liang, Ad hoc mobility management with uniform quorum systems, IEEE Trans. on Networking, Vol.7, No.2, pp.228-240, 1999

[85] L. Blazevic, L. Buttyan, S. Capkun, S. Giordano, J.-P. Hubaux, J-Y. Le Boudec, Self-organization in mobile ad hoc networks: the approach of terminodes, IEEE Comm. Magazine, pp.166-175, June 2001

[86] I. Stojmenovic, Home Agent Based Location Update and Destination Search Schemes in ad hoc Wireless Networks, Tech. rep. TR-99-10, Comp. Science, SITE, Unive. Ottawa, Sept. 1999

[87] F. Kamoun and L. Kleinrock, Hierarchical routing for large networks: Performance evaluation and optimization, Comput. Networks, vol. 1, pp. 155–174, 1977

[88] F. Kamoun and L. Kleinrock, Stochastic performance evaluation of hierarchical routing for large networks, Comput. Networks, vol. 3, pp. 337–353, 1979

[89] J. Sucec, I., Marsic, Hierarchical Routing Overhead in Mobile ad hoc Networks, IEEE Trans. on Mobile Computing, Vol.3, No.1, 2004

[90] R. Perlman, Interconnections: Bridges and Routers. Reading, MA: Addison-Wesley, pp. 149–152, 205–233, 1992

[91] T. Cormen, C. Leiserson, R. Rivest, Introduction to Algorithms, MIT Press, Cambridge MA, 1990

[92] S. Ni, Y. Tseng, Y. Chen, J. Sheu. The broadcast storm problem in a mobile ad hoc network. Proceedings of the ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM), pp. 151–162, 1999.

[93] H. Lim, C. Kim. Multicast tree construction and flooding in wireless ad hoc networks, Proceedings of the ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWIM), pp.61-68, 2000.

[94] R. Sivakumar et al, CEDAR: A Core-Extraction Distributed ad hoc Routing Algorithm, IEEE JSAC, pp. 1454-1465, Aug. 1999

[95] W. Peng, X. Lu. On the reduction of broadcast redundancy in mobile ad hoc networks. Proceedings of MOBIHOC, pp.111-123, 2000.

[96] A. Kershenbaum, Telecommunications network design algorithms, Mcgraw-Hill international editions, pp 282-285, 1993

[97] A. Colvin, CSMA with collision avoidance, Computer Commun., Vol.6, No. 5, pp.227–235, Oct. 1983

[98] S.Xu, T. Saadawi, Does the IEEE 802.11 MAC Protocol Work Well in Multihop Wireless ad hoc Networks?, IEEE Communications Magazine, Vol.39, No.6, pp.130 - 137, June 2001

[99] Z. J. Hass and J. Deng, Dual busy tone multiple access (DBTMA): Performance evaluation, in Proc. 49th VTC, Vol.1, pp.314-319, Oct. 1998

[100] Z. J. Hass and J. Deng, Dual busy tone multiple access (DBTMA)-a multiple access control scheme for ad hoc networks, IEEE Trans. Commun., Vol.50, No.6, pp.975-984, 2002

[101] P. Karn, MACA-A new channel access method for packet radio, in ARRL/CRRL Amateur Radio 9th Computer Networking Conf., pp.134–140, 1990

[102] F. Talucci and M. Gerla, MACA-BI(MACA by invitation): A Wireless MAC Protocol for High Speed Ad hoc Networking, Proc. IEEE ICUPC '97, Vol.2, pp.913-917, 1997

[103] C. L. Fullmer, J. J. Garcia-Luna-Aceves, Solutions to hidden terminal problems in wireless networks, in Proc. ACM SIGCOMM'97, pp.39–49, 1997

[104] J. J. Garcia-Luna-Aceves, C. L. Fullmer, Floor acquisition multiple access (FAMA) in single-channel wireless networks, Mobile Networks and Applications, Vol.4, No.3, pp.157-174, 1999

[105] C.-K. Toh, ad hoc Mobile Wireless Networks: Protocols and Systems, Prentice Hall PTR, Chapter 4, pp.39-56, 2002

[106] S. Wu, et al., Intelligent medium access for mobile ad hoc networks with busy tones and power control, IEEE Journal on Select Areas in Commun., Vol.18, pp.1647-1657, 2000

[107] K. Xu, M. Gerla, S. Bae, Effectiveness of RTS/CTS handshake in IEEE 802.11 based ad hoc networks, Ad Hoc Networks, Vol.1, No.1, pp.107-123, 2003

[108] T. Tsai, C. Tu, An Adaptive IEEE 802.11 MAC in Multihop Wireless ad hoc Networks Considering Large Interference Range, Proceeding of Wireless On-Demand Network Systems 2004, pp.87-100, 2004

[109] K. Xu, M. Gerla, S. Bae, How effective is the IEEE 802.11 RTS/CTS handshake in ad hoc networks? GLOBECOM 2002, No.1, pp.72-77, 2002

[110] J. Li, C. Blake, D. Couto, H. Lee, and R. Morris, Capacity of ad hoc wireless networks, Proceedings of ACM MobiCom'01, pp.61-69, 2001

[111] F. Ye, S. Yi, B. Sikdar, Improving Spatial Reuse of IEEE 802.11 Based ad hoc Networks, Proceedings of IEEE GLOBECOM, pp.1013-1017, 2003

[112] T. Rappaport, Wireless Communications: Principles and Practice, Prentice Hall, Englewood Cliffs, NJ, 1996

[113] M. Cesana, D. Maniezzo, P. Bergamo, M. Gerla, Interference Aware (IA) MAC: an Enhancement to IEEE802.11b DCF, proceedings of the IEEE VTC, Vol.5, pp.2799-2803, fall 2003

[114] H. Takagi, L. Kleinrock, Optimal Transmission Range for Randomly Distributed Packet Radio Terminals, IEEE Trans. Commun., Vol.32, No.3, pp.246–57, 1984

[115] L.Wu and P. Varshney, Performance Analysis of CSMA and BTMA Protocols in Multihop Networks (I). Single Channel Case, Information Sciences, Elsevier Sciences, vol. 120, pp. 159–77, 1999

[116] Y. Wang and J. J. Garcia-Luna-Aceves, Performance of Collision Avoidance Protocols in Single-Channel ad hoc Networks, in Proc. of IEEE Intl. Conf. on Network Protocols, pp.68-77, Nov. 2002

[117] Y. Wang and J. J. Garcia-Luna-Aceves, Modeling of Collision Avoidance Protocols in Single-Channel Multihop Wireless Networks, ACM Wireless Networks, Vol.10, No.5, pp.495-506, 2004

[118] F. Eshghi, A. K. Elhakeem, Performance Analysis of ad hoc Wireless LANs for Real-Time Traffic, IEEE Journal on Select Areas in Comm., Vol.21, No.2, pp.204-215, 2003

[119] G. Bianchi, Performance Analysis of the IEEE 802.11 Distributed Coordination Function, IEEE Journal on Select Areas in Comm., pp.535-547, Vol.18, No.3, 2000

[120] A. Papoulis, S. U. Pillai, Probability, Random Variables and Stochastic Processes, 4th Edition, McGraw-Hill Press, pp.199-201, 243-244, 2002

Appendix A Abbreviations

ACK	ACKnowledgment
AMAC	Adaptive IEEE 802.11 MAC protocol
AODV	Ad hoc On-demand Distance Vector Routing protocol
BTMA	Busy Tone Multiple Access protocol
CBKC	Connectivity Based k -hop Clustering scheme
CCR	Conservative CTS Reply MAC scheme
CDF	Cumulated Density Function
CGSR	Cluster-leader-Gateway Switch Routing protocol
CSMA/CA	Carrier Sense Multiple Access/Collision Avoidance
CoA	Care-of Address
CS	Cluster State packet (or item)
CTS	Clear to Send
DBTMA	Dual Busy Tone Multiple Access protocol
DDCH	Distributed Database Coverage Heuristic scheme
DHCP	Dynamic Host Configuration Protocol
DP	Dominant Pruning broadcasting protocol
DREAM	Distance Routing Effect Algorithm for Mobility routing protocol
DSR	Dynamic Source Routing protocol
FA	Foreign Agent
FAMA	Floor Acquisition Multiple Access protocol
FDMA	Frequency Division Multiple Access
FHS	Forwarding Host Selection broadcast relay scheme
FHSS	Frequency Hopping Spread Spectrum

FPDBT	Fixed Power Dual Busy Tone scheme
GLS	Grid Location Service scheme
GPS	Global Positing System
HA	Home Agent
HLR	Home Location Register
HSR	Hierarchical State Routing protocol
IA	Interference Aware MAC protocol
IEEE 802.11 DCF	IEEE 802.11 Distributed Coordination Function
IntraR	Intra-cluster Routing table
KCBL	k -hop Clustering Based Location Service protocol
KCMBC	k -hop Compound Metric Based Clustering scheme
LC	Local Connectivity table
LLC	Local Link Change packet
LS	Location Service table
LSR	Link State Routing protocol
MAC	Medium Access Control
MACA	Multiple Access Collision Avoidance
MACA-BI	MACA By Invitation
MANET	Mobile Ad Hoc Networks
Max-Min	Max-Min heuristic clustering scheme
MBC	Mobility-Based Clustering scheme
MSC	Minimum Set Covering problem
NAV	Network Allocation Vector
OSFP	Shortest Path First Protocol
PDF	Probability Density Function

PRNET	Multihop multiple access Packet Radio Network
QoS	Quality of Service
RBF	Receiving Beam Forming MAC scheme
RE	Relative Epoch
RTS	Request to Send
TDMA	Time Division Multiple Access
TOA	Time of Arrival
TTL	Time-To-Live
VCS	Virtual Carrier Sensing
VLSI	Very Large Scale Integration
VPDBT	Variable Power Dual Busy Tone scheme
WLAN	Wireless Local Area Network
ZHLS	Zone-based Hierarchical Link State routing protocol
ZRP	Zone Routing Protocol

Appendix B The PDF of the Relative Moving Rate

It is assumed that the moving rate (v_i and v_j) and direction (θ_i and θ_j) of two adjacent mobile hosts h_i and h_j are uniformly distributed among the ranges of $[0, V_m]$ and $[0, 2\pi]$, respectively. The relative moving rate between h_i and h_j is given by:

$$v_d = |\mathbf{v}_d| = |\mathbf{v}_j - \mathbf{v}_i| = \sqrt{v_i^2 + v_j^2 - 2v_i v_i \cos(\theta_i - \theta_j)},$$

Set auxiliary variables: $\alpha = \theta_i - \theta_j$ and $g = cos\alpha$, then

$$v_d = \sqrt{v_i^2 + v_j^2 - 2v_i v_i g}.$$

Since the PDF of θ_i or θ_j is

$$f_{\Theta}(\theta)=\begin{cases}\dfrac{1}{2\pi}, 0\le\theta\le 2\pi\\ 0, \text{ elsewhere}\end{cases},$$

The PDF of α can be obtained as

$$f_{\mathrm{A}}(\alpha)=\int_{-\infty}^{\infty} f_{\Theta}(\alpha+x)f_{\Theta}(x)dx=\begin{cases}\dfrac{\alpha}{4\pi^2}+\dfrac{1}{2\pi}, -2\pi\le\alpha<0\\ -\dfrac{\alpha}{4\pi^2}+\dfrac{1}{2\pi}, 0\le\alpha\le 2\pi\\ 0, \text{ elsewhere}\end{cases} \tag{71}$$

Considering that $g=cos\alpha,\ -2\pi\le\alpha\le 2\pi$,

the solutions are given by

$$\begin{cases}\alpha_1=\arccos g,\ 0\le\alpha_1\le\pi\\ \alpha_2=-\arccos g,\ -\pi\le\alpha_2<0\\ \alpha_3=2\pi-\arccos g,\ \pi<\alpha_3\le 2\pi\\ \alpha_4=\arccos g-2\pi,\ -2\pi\le\alpha_3<-\pi\end{cases}, \tag{72}$$

Then, the PDF of g can be given by

$$f_G(g)=\sum_{i=1}^{4}\frac{f_{\mathrm{A}}(\alpha_i)}{|sin\alpha_i|}, \text{ for } -1<g<1$$

Substitute (71) and (72) into the above equation:

$$f_G(g)=\begin{cases}\dfrac{1}{\pi\sqrt{1-g^2}}, -1<g<1\\ 0, \text{ elsewhere}\end{cases} \tag{73}$$

Let $\begin{cases}v_d=\sqrt{v_i^2+v_j^2-2v_iv_ig}\\ x=v_i\\ y=v_j\end{cases}$, then the solution is $\begin{cases}g=-\dfrac{v_d^2-x^2-y^2}{2xy}\\ v_i=x\\ v_j=y\end{cases}$

The Jacobian transformation is given by

$$J(v_i, v_j, g) = \begin{vmatrix} \frac{\partial v_d}{\partial v_i} & \frac{\partial v_d}{\partial v_j} & \frac{\partial v_d}{\partial g} \\ \frac{\partial x}{\partial v_i} & \frac{\partial x}{\partial v_j} & \frac{\partial x}{\partial g} \\ \frac{\partial y}{\partial v_i} & \frac{\partial y}{\partial v_j} & \frac{\partial y}{\partial g} \end{vmatrix} = \begin{vmatrix} \frac{\partial v_d}{\partial v_i} & \frac{\partial v_d}{\partial v_j} & \frac{\partial v_d}{\partial g} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = -\frac{xy}{v_d}$$

According to [120], the joint density of v_d, x and y can be obtained as

$$f_{V_d XY}(v_d, x, y) = \frac{f_{V_i V_j G}(v_i, v_j, g)}{\left|J(v_i, v_j, g)\right|}$$

Since v_i, v_j and g are independent variables, then

$$f_{V_i V_j G}(v_i, v_j, g) = f_V(v_i) f_V(v_j) f_G(g),$$

Considering equation (73) and $f_V(v) = \begin{cases} \frac{1}{V_m}, 0 \le v \le V_m \\ 0, \text{ elsewhere} \end{cases}$, then

$$f_{V_d XY}(v_d, x, y) = \frac{f_{V_i V_j G}(v_i, v_j, g)}{\left|J(v_i, v_j, g)\right|} = \frac{2v_d}{\pi V_m^2 \sqrt{4x^2 y^2 - (v_d^2 - x^2 - y^2)^2}}$$

The PDF of v_d can be calculated as

$$f_{V_d}(v_d) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{V_d XY}(v_d, x, y) dx dy$$

As $x = v_i \in [0, V_m]$, $y = v_j \in [0, V_m]$ and $g = -\frac{v_d^2 - x^2 - y^2}{2xy}$, $g \in (-1, 1)$, then

$\left|x - y\right| < v_d < x + y$ and $0 \le v_d \le 2V_m$

Consequently, the PDF of the relative moving rate is given by:

$$f_{V_d}(v_d) = \begin{cases} \frac{2v_d}{\pi V_m^2} \iint\limits_{\substack{0 \le x \le V_m \\ 0 \le y \le V_m \\ \left|x-y\right| < v_d < x+y}} \frac{1}{\sqrt{4x^2 y^2 - (v_d^2 - x^2 - y^2)^2}} dx dy, & 0 \le v_d \le 2V_m \\ 0, & \text{elsewhere} \end{cases}$$

Appendix C Correctness of KCMBC

The correctness of KCMBC is solely dependant on hosts electing cluster-heads that actually become cluster-heads. The following Claims are used to show that the hosts, which survive after the Floodmax stage of KCMBC, become cluster-heads.

Claim 1: During the Floodmin and Floodmax algorithms, no host's compound metric will propagate farther than k -hops from the source host itself.

Claim 2: All hosts that survive after Floodmax stage elect themselves as cluster-heads.

Proof: The Floodmax stage generates a dominating set of hosts which consists of two classes of hosts: Class1 hosts are those hosts whose metrics are the largest in their k -hop neighborhood, and Class2 hosts are those hosts whose metrics are the largest in at least one of their k -hop neighbors' k -neighborhood.

Considering a Class1 host h_i, it will overtake each host that is k -hop away during the Floodmax. Hence, all hosts that are within the k -hop distance of h_i will possess h_i's ID in the WINNER data structure. Then, after the Floodmin stage, a Class1 host will elect itself as the cluster-head.

On the other hand, consider a Class2 host h_j. Although h_j is overtaken by larger compound metric, its compound metric continues to propagate out and overtake all smaller metrics within k -hop of h_j. Therefore, at the end of the Floodmax stage, h_j's compound metric and larger compound metrics will cover the k -hop distance of h_j. Therefore, the compound

metric of Class2 host h_j is the smallest surviving metric in the k-hop neighborhood of h_j. Based on Claim 1, it can be concluded that the Floodmin process will successfully propagate the compound metric of h_j back to h_j. A Class2 host will elect itself as the cluster-head, based on *Rule 1*. Therefore, any host that survives the Floodmax stage will elect itself as a cluster-head.

Based on the above Claims, the following Lemma shows that every host that is elected as a cluster-head does become a cluster-head. Consequently, the correctness of KCMBC can be proved.

Lemma: If host h_i elects host h_j as its cluster-head, then host h_j becomes a cluster-head.

Proof: According to the three rules for cluster-head election shown in Section 3.3.2, the election of cluster-heads can be considered as the following three cases.

Case 1: Host h_i elects itself as a cluster-head based on *Rule 1*. If h_i receives its own compound metric in the Floodmin stage, it knows that other hosts have elected it as cluster-head based on Claim 2. Therefore, it elects itself a cluster-head.

Case 2: Host h_i elects host h_j as its cluster-head based on *Rule 2*. In this case, h_i receives h_j's compound metric in the Floodmin stage. Hence, it can be concluded that h_j does become a cluster-head based on Claim 2.

Case 3: Host h_i elects host h_j as its cluster-head based on *Rule 3*. In this case, h_i must select the host h_j, which is the last WINNER of the Floodmax, to be a cluster-head. This host h_j survives the Floodmax and become a cluster-head based on Claim 2.

Appendix D The Cluster Parameters for KCMBC

Table A- 1. The average number (N_h) of members in a cluster

N / k	1000	2000	3000	4000
1	2.65	3.39	4.86	5.54
2	3.89	7.35	11.54	15.87
3	4.85	12.55	19.82	26.26
4	5.88	17.39	32.96	50.01
5	6.45	23.42	43.20	66.38

Table A- 2. The average number (d_C) of neighboring clusters connected with a cluster

N / k	1000	2000	3000	4000
1	1.96	4.72	7.37	8.60
2	1.52	4.52	6.65	8.02
3	1.19	4.50	6.40	7.47
4	0.99	4.45	6.24	7.15
5	0.77	4.41	6.18	7.10

Appendix E Area Calculation for MAC Protocols

1. The Crescent Shape Area $\omega(r_A, r_B, x)$

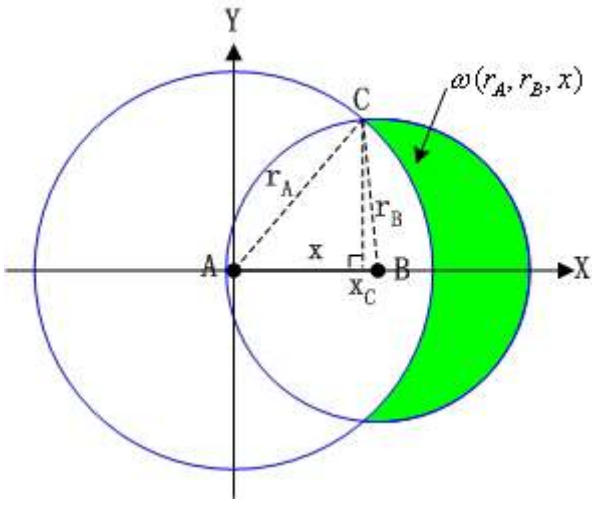


Figure A- 1. The area of the crescent shape $\omega(r_A, r_B, x)$

As shown in Figure A- 1, it is supposed that two circles with radius r_A and r_B ($r_A > r_B$) are centered at point A and B, respectively, and the distance between A and B is x. The shadow area $\omega(r_A, r_B, x)$ is inside of the circle B but outside of the circle A. Let x_C denote the abscissa of the intersection points of the two circumferences, then if $r_A - r_B < x \le r_A$

$$r_A^2 - x_C^2 = r_B^2 - (x - x_C)^2$$

Thus, $x_C = \dfrac{r_A^2 - r_B^2 + x^2}{2x}$,

and $\omega(r_A, r_B, x) = 2\int_{x_C}^{x+r_B} \sqrt{r_B{}^2 - (t-x)^2}\,dt - 2\int_{x_C}^{r_A} \sqrt{r_A{}^2 - x^2}\,dt$, for $r_A - r_B < x \le r_A$

Therefore, the area of the crescent shape $\omega(r_A, r_B, x)$ is given by

$$\omega(r_A, r_B, x) = \begin{cases} 0, & x \le r_A - r_B \\ \dfrac{1}{2}[\pi(r_B{}^2 - r_A{}^2) + \delta(r_A, r_B, x)] + r_A{}^2 \arctan\left[\dfrac{x^2 + r_A{}^2 - r_B{}^2}{\delta(r_A, r_B, x)}\right] & \\ + r_B{}^2 \arctan\left[\dfrac{x^2 - r_A{}^2 + r_B{}^2}{\delta(r_A, r_B, x)}\right], & r_A - r_B < x \le r_A \end{cases}$$

where $\delta(r_A, r_B, x) = \sqrt{4x^2 r_A{}^2 - (x^2 + r_A{}^2 - r_B{}^2)^2}$

2. Interfering Areas of IEEE 802.11

Considering that host i intends to transmit a data packet to host j which is at the distance x, let r, r_I and r_{Imax} denote the radius of the transmission range, the interference range and the maximum the interference range of host i or j, respectively. Based on the range of the distance x, the relationship between the transmission range and the interference range can be categorized into three cases which are shown in Figure

A- 2. Using the area notations defined in Section 6.3, the interfering areas for each case, which are listed in Table 6-1, are derived as follows.

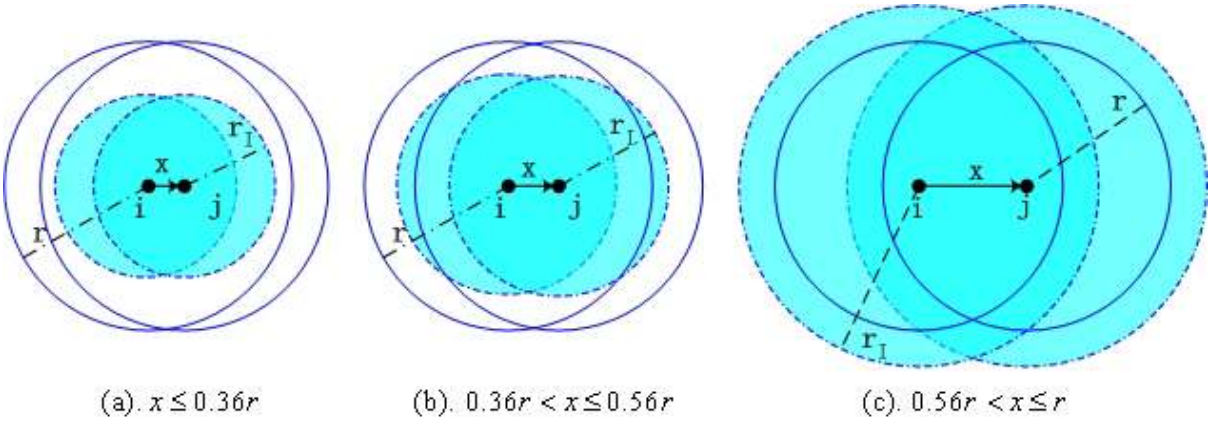


Figure A- 2. Three cases of interfering area

(a). $x \le 0.36r$

As shown in Figure A- 2(a), in this case, the transmission range of host i or j can totally cover the interference ranges of the communication pair. Since all the hosts located inside of the interference ranges of hosts i and j can hear the RTS/CTS handshake, no packet will be transmitted by these hosts during the transmissions of CTS, data and ACK packets. The possible collisions happen only if a host in j's interference range transmits a packet at the time slot when host i transmit RTS. Hence, none of hosts in the interfering area $S_{I_{ij}}(x) = \pi r_I^2$ should transmit for 1 slot.

(b). $0.36r \le x \le 0.56r$

Figure A- 2(b) shows that when $0.36r \le x \le 0.56r$, the transmission range is larger than the interference range, but the transmission range of host i or j cannot cover all the interference ranges of both host i and host j. As the RTS/CTS handshake can be heard

by all hosts located inside of the interference range, collisions can only happen during the transmission of RTS.

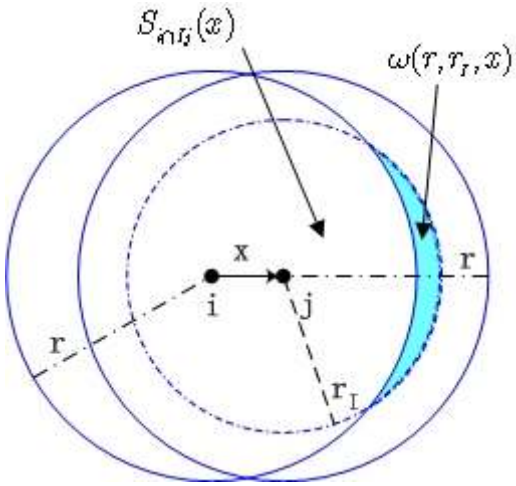


Figure A- 3. The interfering area for $0.36r \le x \le 0.56r$

As shown in Figure A- 3, in order to avoid collisions when host i transmits a RTS to host j, the hosts located inside of the area $S_{i\cap Ij}(x)$, which is inside j's interference range and covered by i's transmission range, must defer their transmissions for 1 slot. The interfering area $S_{i\cap Ij}(x)$ is given by

$$S_{i\cap Ij}(x) = S_{Ij}(x) - S_{Ij-i}(x) = \pi r_I^2 - \omega(r, r_I, x),$$

where $S_{Ij-i}(x) = \omega(r, r_I, x)$ is the shadow area shown in Figure A- 3, and the expression of $\omega(r_A, r_B, x)$ can be found in Appendix E.1. Moreover, since the hosts located inside of the shadow area $S_{Ij-i}(x) = \omega(r, r_I, x)$ cannot hear RTS, none of hosts in this area should transmit for $2t_{RTS} + 1$ slots.

(c). $0.56r \le x \le r$

The case of $0.56r \le x \le r$ is shown in Figure A- 2(c), where the interference range of host i or j totally contains its transmission range, i.e., $r_I > r$. In this case, Collision may happen in each step of the four-way handshake of IEEE 802.11.

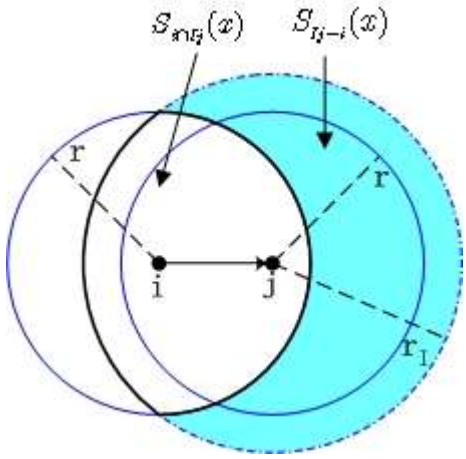


Figure A- 4. The interfering area for RTS transmission, $0.56r \le x \le r$

When host i transmits a RTS to host j, the hosts located inside of the area $S_{i\cap Ij}(x)$, which is covered by i's transmission range and inside j's interference range, can hear RTS. Hence, these hosts cannot transmit for 1 slot. As shown in Figure A- 4, $S_{i\cap Ij}(x)$ is the area enveloped by the bold arcs, which can be given by

$$S_{i\cap Ij}(x) = S_i(x) - S_{i-Ij}(x) = \pi r^2 - \omega(r_I, r, x).$$

On the other hand, since the RTS packet cannot reach the shadow area $S_{Ij-i}(x)$, which is inside of j's interference range but outside of i's transmission range, no host in the area $S_{Ij-i}(x)$ is allow to transmit for $2t_{RTS}+1$ slots. $S_{Ij-i}(x)$ can be obtained as

$$S_{Ij-i}(x) = S_{Ij}(x) - S_{i\cap Ij}(x) = \pi(r_I^2 - r^2) + \omega(r_I, r, x)$$

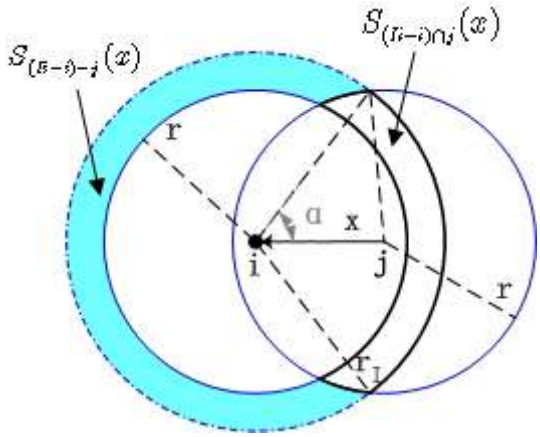


Figure A- 5. The interfering area for CTS transmission, $0.56r \le x \le r$

Upon receiving a correct RTS, host j replies a CTS to i. The hidden hosts which are located inside of i's interference range but outside of i's transmission range may interrupt the reception of CTS at host i. Among the hidden hosts, those which are also located inside of j's transmission range can hear CTS, so they need to defer transmissions for 1 slot. As shown in Figure A- 5, the interfering area for such hidden hosts, denoted as $S_{(Ii-i)\cap j}(x)$, is the area enveloped by the bold arcs. $S_{(Ii-i)\cap j}(x)$ can be approximated as

$$S_{(Ii-i)\cap j}(x) \approx \frac{2\alpha(x)}{2\pi} \times \pi(r_I^2 - r^2) = \alpha(x)(r_I^2 - r^2)$$

where the angle $\alpha(x) = \arccos\dfrac{r_I^2 + x^2 - r^2}{2r_I x}$.

On the other hand, the hidden hosts, which are located inside of the shadow area $S_{(Ii-i)-j}(x)$ shown in Figure A- 5 (i.e. the area outside of the transmission ranges of both i and j), cannot transmit for $t_{RTS} + t_{CTS} + 1$ slots. The interfering area $S_{(Ii-i)-j}(x)$ can be given by

$$S_{(Ii-i)-j}(x) = S_{Ii}(x) - S_i(x) - S_{(Ii-i)\cap j}(x) \approx [\pi - \alpha(x)](r_I^2 - r^2)$$

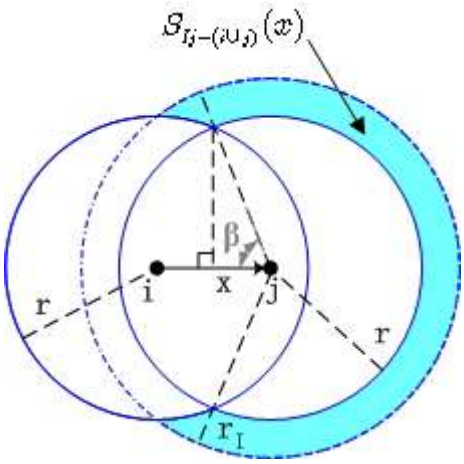


Figure A- 6. The interfering area for data transmission, $0.56r \le x \le r$

If the RTS/CTS handshake is successful, the hosts located inside of the transmission ranges of hosts i and j will defer their transmission when host i transmits a data packet. However, the hidden hosts, which cannot hear the RTS/CTS handshake but are located inside of j's interference range, may interrupt the reception of the data packet. The vulnerable period is $t_{RTS} + t_{DATA} + 1$ slots. As shown in Figure A- 6, the interfering area for these hidden hosts is the shadow area $S_{Ij-(i\cup j)}(x)$. In Figure A- 6, the degree β is obtained as

$$\beta = \arccos[x/(2r)]$$

Then, $S_{Ij-(i\cup j)}(x)$ can be approximated by:

$$S_{Ij-(i\cup j)}(x) \approx \pi(r_I^2 - r^2) \times \frac{2\beta}{2\pi} = (r_I^2 - r^2) \times \{\pi - \arccos[x/(2r)]\}$$

Likewise, when host j replies host i an ACK to confirm a successful data transmission, none of hosts located inside of the interfering area $S_{Ii-(i\cup j)}(x)$ should transmit for $t_{RTS} + t_{ACK} + 1$ slots, where $S_{Ii-(i\cup j)}(x)$ is approximated as

$$S_{Ii-(i\cup j)}(x) \approx (r_I^2 - r^2) \times \{\pi - \arccos[x/(2r)]\}.$$