Technical Report (Nr. 2003-08)

A Survey of Mobile Ad Hoc network Routing Protocols*

Changling Liu, Jörg Kaiser

{changling.liu, joerg kaiser}@informatik.uni-ulm.de

Department of Computer Structures
University of Ulm, Germany

October 2003

* The work was partially supported by the EC, through project IST-2000-26031 (CORTEX - CO-operating Real-time senTient objects: architecture and EXperimental evaluation).

Abstract

Emerging ubiquitous computing needs "anytime and anywhere" network connections. Mobile ad hoc networks are well suited for this application area because they are self-organizing networks without depending on any network infrastructure. Due to the improved flexibility and reduced cost of mobile ad hoc networks, they promise to play an important role in future mobile applications. An important and essential issue for mobile ad hoc networks is routing protocol design. Because of the dynamic network features, it is a major technical challenge. During the last years, active research work resulted in a variety of proposals. A large number of protocols, each with a particular property and often optimized for a specific application area, have been designed. They follow different design principles and exhibit substantial variations in performance depending on network size or node mobility patterns. This report presents a state-of-the-art review and a comparison for typical representatives of routing protocols designed for mobile ad hoc networks. The report aims at providing criteria according to which the plethora of protocols can be classified. Because it is difficult to compare all protocols along the same criteria, which some protocols may have others not, we first identify major protocol classes and then provide an in depth comparison of related protocols of the same category.

Contents

1 Introduction

With the advances of wireless communication technology, low-cost and powerful wireless transceivers are widely used in mobile applications. Mobile networks have attracted significant interests in recent years because of their improved flexibility and reduced costs.

Compared to wired networks, mobile networks have unique characteristics. In mobile networks, node mobility may cause frequent network topology changes, but such topology changes are thought as rare exceptions in wired networks. In contrast to the stable link capacity of wired networks, wireless link capacity continually varies because of the impacts from transmission power, receiver sensitivity, noise, fading and interference. Additionally, wireless mobile networks have a high error rate, power restrictions and bandwidth limitations.

Mobile networks can be classified into infrastructure networks and mobile ad hoc networks [1] according to their dependence on fixed infrastructures. In an infrastructure mobile network, mobile nodes have wired access points (or base stations) within their transmission range. The access points compose the backbone for an infrastructure network. In contrast, mobile ad hoc networks are autonomously self-organized networks without infrastructure support. In a mobile ad hoc network, nodes move arbitrarily, therefore the network may experiences rapidly and unpredictably topology changes. Additionally, because nodes in a mobile ad hoc network normally have limited transmission ranges, some nodes cannot communicate directly with each other. Hence, routing paths in mobile ad hoc networks potentially contain multiple hops, and every node in mobile ad hoc networks has the responsibility to act as a router.

Mobile ad hoc networks originated from the DARPA Packet Radio Network (PRNet) [35] and SURAN project [36]. Being independent on pre-established infrastructure, mobile ad hoc networks have advantages such as rapid and ease of deployment, improved flexibility and reduced costs. Mobile ad hoc networks are appropriate for mobile applications either in hostile environments where no infrastructure is available, or temporarily established mobile applications which are cost crucial. In recent years, application domains of mobile ad hoc networks have been expanded civilian and industrial areas. The typical application scenarios include the rescue missions, the law enforcement operations, the cooperating industrial robots, the traffic management, and the educational operations in campus.

Active research work for mobile ad hoc network is carrying on mainly in the fields of medium access control, routing, resource management, power control and security. Because of the importance of routing protocols in dynamic multi-hop networks, a lot of mobile ad hoc network routing protocols have been proposed in the last few years. There are some challenges that make the design of mobile ad hoc network routing protocols a tough task. Firstly, in mobile ad hoc networks, node mobility causes frequent topology changes and network partitions. Secondly, because of the variable and unpredictable capacity of wireless links, packet losses may happen frequently. Moreover, the broadcast nature of wireless medium introduces the hidden terminal and exposed terminal problems. Additionally, mobile nodes have restricted power, computing and bandwidth resources and require effective routing schemes.

As a promising network type in future mobile applications, mobile ad hoc networks are attracting more and more researchers. This report gives the state-of-the-art review for typical routing protocols for mobile ad hoc networks, including classical MANET unicast and multicast routing algorithms and popular classification methods. In this report, related routing protocols are compared from an analysis point of view based on the classification methods.

The remainder of this report is organized as follows. Section 2 gives a short introduction of routing algorithms used in wired networks and the reasons why they are not suitable for mobile ad hoc networks. In Section 3, evaluation methods for mobile ad hoc routing protocols are given. Several classification methods are given. 4. In section 5, typical unicast and multicast routing protocols for mobile ad hoc networks are reviewed. Analysis and a comparison for different routing approaches are also given in this section. Section 6 concludes the report.

2 Limitations of traditional routing approaches

Routing is a fundamental issue for networks. A lot of routing algorithms have been proposed for wired networks and some of them have been widely used. Nowadays, dynamic routing approaches are prevalent in wired networks. Distance Vector routing [31] and Link State routing [31] are two of the most popular dynamic routing algorithms used in wired networks.

Distance Vector routing algorithm is also called Bellman-Ford routing algorithm. In Distance Vector routing, every router maintains a routing table (i.e. vector), in which it stores the distance information to all reachable destinations. A router exchanges distance information with its neighbors periodically to update its routing table. The distance can be calculated based on metrics such like hop number, queue length or delay. If multiple paths exist, the shortest one will be selected. The main drawback of Distance Vector routing algorithm is the slow convergence. Slow convergence leads to the "count-to-infinity" problem, i.e., some routers continuously increase the hop count to particular networks. Open Shortest Path First (OSPF) [54] is a typical Internet routing protocol based on Distance Vector algorithm.

In Link State routing algorithm, each node periodically notifies its current status of links to all routers in the network. Whenever a link state change occurs, the respective notifications will be flooded throughout the whole network. After receiving the notifications, all routers re-compute their routes according to the fresh topology information. In this way, a router gets to know at least a partial picture of the whole network. In Link State routing, different metrics can be chosen, such like number of hops, link speed and traffic congestion. Shortest (or lowest cost) paths are calculated using Dijkstra's algorithm. Routing Information Protocol (RIP) [55] is an Internet routing protocol based on Link State routing algorithm.

In wired networks, Distance Vector and Link State routing algorithms perform well because of the predictable network properties, such as static link quality and network topology. However, the dynamic features of mobile ad hoc networks deteriorate their effectiveness. In mobile ad hoc networks, when using a Distance Vector routing or Link State based routing protocol designed for wired networks, frequent topology changes will greatly increase the control overhead. Without remedy, the overhead may overuse scarce bandwidth of mobile ad hoc networks. Additionally, Distance Vector and Link State routing algorithms will cause routing information inconsistency and route loops when used for dynamic networks.

Multicast is required by applications in which subsets of nodes have common interests for specific information. In such scenarios, multicast out-performs unicast due to the saving of bandwidth and computing resource. Multicast routing, together with multicast addressing and dynamic registration, provides supports for multicast in wired networks. The multicast routing avoids multiple transmissions of the same message to receivers belonging to the same subset. Many multicast routing schemes have been proposed for wired networks, both Internet and ATM. Multicast routing approaches, such as Distance Vector Multicast Routing Protocol (DVMRP) [50], Multicast Open Shortest Path First (MOSPF) [51], Protocol-Independent Multicast (PIM) [52, 53] have been wildly used in wired networks. In most of them, distributed trees are built from the sender to receivers belonging to same group. DVMRP adopts reverse path forwarding and needs periodically flooding to discover new hosts that want to join a particular group. Some multicast routing protocols are dependent on specific unicast routing protocols. For example, Multicast OSPF (MOSPF) is an extension to the OSPF unicast routing protocol and includes multicast information in link state advertisements. Each MOSPF router knows all multicast groups currently residing in the network and builds a distribution tree for each source/group pair. Distribution trees must be re-computed either periodically or when there is a link state change. In contrast to MOSPF, the Protocol-Independent Multicast (PIM) was proposed to work with all existing unicast routing protocols. There are two types of PIM: the dense-model PIM and the sparse-model PIM. The dense-mode PIM was designed for environments where group members are packed relatively densely and enough bandwidth resource is available. Whereas the sparse-mode PIM refers to environments where group members are distributed across many regions of the network and bandwidth is scarce.

As in wired networks, multicast is also appealing for mobile ad hoc networks. Multicast is an appropriate communication scheme for many mobile applications and can save bandwidth resource of wireless channels. Additionally, the inherent broadcast property of wireless channels can be exploited to improve multicast performance in mobile ad hoc networks. Compared to unicast routing schemes, designing multicast routing protocols for mobile ad hoc networks is more difficult. The node mobility makes keeping track of the multicast group membership more complicated and expensive than in wired networks. Also, a distribution tree suffers from frequent reconstruction because of node movements. Therefore, multicast routing schemes for mobile ad hoc networks must include mechanisms to cope with the difficulties incurred by node mobility and topology changes.

3 Evaluation methods

Although numerous routing protocols have been proposed for mobile ad hoc networks, there is no “one-for-all” scheme that works well in scenarios with different network sizes, traffic overloads, and node mobility patterns. Moreover, those protocols are based on different design philosophies and proposed to meet specific requirements from different application domains. Thus, the performance of a mobile ad hoc routing protocol may vary dramatically with the variations of network status and traffic overhead. The performance variations of mobile ad hoc network routing protocols make it a very difficult task to give a comprehensive performance comparison for a large number of routing protocols.

There are three different ways to evaluate and compare the performance of mobile ad hoc routing protocols. The first one is based on analysis [38] and uses parameters such as time complexity, communication complexity for performance evaluation. In the second method, routing performance is compared based to simulation results [9,39]. Network Simulator [27], GloMoSim [28] and OPNET [29] are wildly used simulators. The simulation results heavily dependent on the selection of simulation tools and configuration of simulation parameters. The last method is implementing routing protocols and analysis their performance using data from real-world implementations. This method is not suitable for comparison of a large number of routing protocols.

Considering the dynamic network features, metrics for evaluating performance of mobile ad hoc network routing protocols are proposed in [2]. Generally, a properly designed mobile ad hoc routing protocol should adapt to the dynamic network changes quickly with lower consumption of communication and computing resource.

4 Classification methods for routing protocols

To compare and analyze mobile ad hoc network routing protocols, appropriate classification methods are extremely important. Classification methods help researchers and designers to understand distinct characteristics of a routing protocol and find its intrinsic relationship with others. There exist different classification methods for mobile ad hoc network routing approaches, which are based on diverse criteria and from specific viewpoints. These classification methods are not completely orthogonal. According to different classification methods, a routing algorithm can be categorized into different groups.

According to the most commonly used classification method, for both wired networks and wireless networks, routing algorithms can be divided into unicast, multicast and broadcast. This report focuses on unicast and multicast routing protocols for mobile ad hoc networks. In this section, classification methods particularly used for mobile ad hoc network routing protocols are given.

4.1 Proactive routing, reactive routing and hybrid routing

The most popular method being used for classifying mobile ad hoc network routing protocols is based on how routing information is acquired and maintained by mobile nodes. Using this method, mobile ad hoc network routing protocols can be divided into proactive routing, reactive routing and hybrid routing.

A proactive routing protocol is also called "table driven" routing protocol. Using a proactive routing protocol, nodes in a mobile ad hoc network continuously evaluate routes to all reachable nodes and attempt to maintain consistent, up-to-date routing information. Therefore, a source node can get a routing path immediately if it needs one.

In proactive routing protocols, all nodes need to maintain a consistent view of the network topology. When a network topology change occurs, respective updates must be propagated throughout the network to notify the change. Most proactive routing protocols proposed for mobile ad hoc networks have inherited properties from algorithms used in wired networks. To adapt to the dynamic features of mobile ad hoc networks, necessary modifications have been made on traditional wired network routing protocols. Using proactive routing algorithms, mobile nodes proactively update network state and maintain a route regardless of whether data traffic exists or not, the overhead to maintain up-to-date network topology information is high. In Section 4, we will give introductions of several typical proactive mobile ad hoc network routing protocols, such as the Wireless Routing Protocol (WRP) [32], the Destination Sequence Distance Vector (DSDV) [4] and the Fisheye State Routing (FSR).

Reactive routing protocols for mobile ad hoc networks are also called "on-demand" routing protocols. In a reactive routing protocol, routing paths are searched only when needed. A route discovery operation invokes a route-determination procedure. The discovery procedure terminates either when a route has been found or no route available after examination for all route permutations.

In a mobile ad hoc network, active routes may be disconnected due to node mobility. Therefore, route maintenance is an important operation of reactive routing protocols. Compared to the proactive routing protocols for mobile ad hoc networks, less control overhead is a distinct advantage of the reactive routing protocols. Thus, reactive routing protocols have better scalability than proactive routing protocols in mobile ad hoc networks. However, when using reactive routing protocols, source nodes may suffer from long delays for route searching before they can forward data packets. The Dynamic Source Routing (DSR) [5] and Ad hoc On- demand Distance Vector routing (AODV) [17] are examples for reactive routing protocols for mobile ad hoc networks.

Hybrid routing protocols are proposed to combine the merits of both proactive and reactive routing protocols and overcome their shortcomings. Normally, hybrid routing protocols for mobile ad hoc networks exploit hierarchical network architectures. Proper proactive routing approach and reactive routing approach are exploited in different hierarchical levels, respectively. In this report, as examples of hybrid routing protocols for mobile ad hoc networks, the Zone Routing Protocol (ZRP) [8], Zone-based Hierarchical Link State routing (ZHLS) [10] and Hybrid Ad hoc Routing Protocol (HARP) [11] will be introduced and analyzed.

4.2 Uniform routing and non-uniform routing

Another classification method for mobile ad hoc network routing protocols is based on the functional differences between mobile nodes. Using this criterion, existing mobile ad hoc network routing protocols can be divided into uniform routing and non-uniform routing.

In a uniform routing protocol for mobile ad hoc networks, all mobile nodes have same importance and functionality. Examples of uniform routing protocols include Wireless Routing Protocol (WRP), Dynamic Source Routing (DSR), Ad hoc On-demand Distance Vector routing (AODV) and Destination Sequence Distance Vector (DSDV) routing protocol. Uniform routing protocols normally assume a flat network structure.

In a non-uniform routing protocol for mobile ad hoc networks, some nodes carry out distinct management and/or routing functions. Normally, distributed algorithms are exploited to select those special nodes. In some cases, non-uniform routing approaches are related to hierarchical network structures to facilitate node organization and management.

Non-uniform mobile ad hoc network routing protocols can be divided further according to both how the way to organize organization manners of mobile nodes and how management and routing functions are performed. Following these criteria, non-uniform routing protocols for mobile ad hoc networks are divided into zone based hierarchical routing; cluster-based hierarchical routing and core-node based routing.

In zone based routing protocols, different zone constructing algorithms are exploited for node organization. Some zone constructing algorithms uses geographical information. Some zone constructing methods result overlapping zones and some don’t. Exploiting zone division effectively reduces the overhead for routing information maintenance. Mobile nodes in the same zone know how to reach each other with smaller cost compared to maintaining routing information for all nodes in the whole network. In some zone based routing protocols, specific nodes act as gateway nodes and carry out inter-zone communication. The Zone Routing Protocol (ZRP) and Zone-based Hierarchical Link State routing (ZHLS) [10] are zone based hierarchical routing protocols for mobile ad hoc networks.

A cluster based routing protocol uses specific clustering algorithm for clusterhead election. Mobile nodes are grouped into clusters and clusterheads take the responsibility for membership management and routing functions. Clusterhead Gateway Switch Routing (CGSR) [13] will be introduced in Section 5 as an example of cluster based mobile ad hoc network routing protocols. Some cluster based mobile ad hoc network routing protocols potentially support a multi-level cluster structure, such as the Hierarchical State Routing (HSR) [23].

In core-node based routing protocols for mobile ad hoc networks, critical nodes are dynamically selected to compose a "backbone" for the network. The “backbone” nodes carry out special functions, such as routing paths construction and control/data packets propagation. Core-Extraction Distributed Ad Hoc Routing (CEDAR) [22] is a typical core-node based mobile ad hoc network routing protocols.

4.3 Routing metrics based classification

Metrics used for routing path construction can be used as criteria for mobile ad hoc network routing protocol classification. Most routing protocols for mobile ad hoc networks use "hop number" as metric. If there are multiple routing paths available, the path with minimum hop number will be selected. If all wireless links in the network have the same failure probability, short routing paths are more stable than the long ones and can obviously decrease traffic overhead and reduce packet collisions. However, this assumption is not true in mobile ad hoc networks.

Selecting stable links during the route path construction phase can ease the maintenance overhead in mobile ad hoc networks. For example, routing approaches such as Associativity Based Routing (ABR) [24] and Signal Stability based Routing (SSR) [26] are proposed that use link stability or signal strength as metric for routing.

With the popularity of mobile computing, some mobile applications may have different QoS requirements. To meet specific QoS requirements, appropriate QoS metrics should be used for packet routing and forwarding in mobile ad hoc networks. As in wired networks, QoS routing protocols for mobile ad hoc networks can use metrics, such as bandwidth, delay, delay jitter, packet loss rate and cost. As an example, bandwidth and link stability are used in CEDAR as metrics for routing path construction.

4.4 Topology based routing and destination based routing

In a topology based routing protocol for mobile ad hoc networks, nodes collect network topology information for making routing decisions. Other than topology based routing protocols, there are some destination-based routing protocols proposed in mobile ad hoc networks. In a destination–based routing protocol, when forwarding a packet to the destination, a node only need to know the next hop along the routing path. For example, DSR is a topology based routing protocol. AODV and DSDV are destination based routing protocols.

4.5 Location based routing and content based routing

Nowadays, availability of GPS or similar locating systems allows mobile nodes to access geographical information easily. In location-based routing protocols, the position relationship between a packet forwarding node and the destination, together with the node mobility can be used in both route discovery and packet forwarding. Existing location-based routing approaches for mobile ad hoc networks can be divided into two schemes. In the first scheme, mobile nodes send packets merely depending on the location information and do not need any extra knowledge. The other scheme uses both location information and topology information. Location Aided Routing (LAR) [6] and Distance Routing Effect Algorithm for Mobility (DREAM) [15] are typical location-based routing protocols proposed for mobile ad hoc networks.

For some mobile applications in mobile ad hoc networks, the content-based paradigm is more suitable for autonomous communication scheme. The Content Based Multicast (CBM) [21] is an example of content-based mobile ad hoc network routing protocol.

4.6 Classify multicast routing protocols for mobile ad hoc networks

Most classification methods used for unicast routing protocols for mobile ad hoc networks are also applicable for existing multicast routing protocols. For example, multicast routing algorithms for mobile ad hoc networks can be classified into reactive routing and proactive routing. The Ad-hoc Multicast Routing (AMRoute) [46] and Ad hoc Multicast Routing protocol utilizing Increasing id-numberS (AMRIS) [47] belong to category of proactive multicast routing and the On-Demand Multicast Routing Protocol *(*ODMRP) [44] and Multicast Ad hoc On-demand Distance Vector (MAODV) [61] are reactive multicast routing protocols.

There is a classification method particularly used for multicast routing protocols for mobile ad hoc networks. This method is based on how distribution paths among group members are constructed. According to this method, existing multicast routing approaches for mobile ad hoc networks can be divided into tree based multicast routing, mesh based multicast routing, core based multicast routing and group forwarding based multicast.

Tree based multicast routing protocols can be further divided into source-rooted and core-rooted according to roots of the multicast trees. In source-rooted tree based multicast routing protocols, source nodes are roots of multicast trees and execute algorithm for distribution tree contraction and maintenance. This requires that a source must know the topology information and addresses of all its receivers in the multicast group. Therefore, source-rooted tree based multicast routing protocols suffer from control traffic overhead when used for dynamic networks. The AMRoute [46] is an example for source-rooted tree multicast routing.

In a core-based multicast routing protocol, cores are nodes with special functions such as multicast data distribution and membership management. Some core-based multicast routing protocols utilize tree structures also, but unlike source-rooted tree based multicast routing, multicast trees are rooted at core nodes. For different core-based multicast routing protocols, core nodes may perform various routing and management functions. For example, in CTB [60] and AMRIS, cores are cross points for all traffic flows of multicast groups and may becomes bottlenecks of the network. On the other hand, in protocols like CAMP, core nodes are not necessarily to be parts of all routing paths.

In a mesh-based multicast routing protocol, packets are distributed along mesh structures that are a set of interconnected nodes. The mesh structure is more robust than the tree structure when used for multicast routing in dynamic networks because a mesh provides alternate paths when link failure occurs. However, the cost for maintaining mesh structures are normally higher than trees. The ODMRP [44] and Core-Assisted Mesh Protocol (CAMP) [45] are mesh-based multicast routing protocols proposed for mobile ad hoc networks.

In the group forwarding based multicast routing, a set of mobile nodes is dynamically selected as forwarding nodes for a multicast group. Forwarding nodes take the responsibility for multicast packet distribution. Using this scheme, it is possible to get multiple routing paths, and duplicate messages will reach a receiver through different paths. ODMRP is a group forwarding based multicast routing protocol using adaptive forwarding groups.

5 Analysis and comparison for typical routing protocols

Because lots of routing protocols have been proposed for mobile ad hoc networks, it is impossible to cover all of them in this review. Therefore, this report presents some typical protocols that can reflect the state-of-the-art of research work on mobile ad hoc network routing. The protocols are arranged according to their classifications. Table 1 and 2 list the protocols reviewed in this report.

Table 1. Unicast routing protocols reviewed in this report

Uniform routing	Proactive routing	Wireless Routing Protocol (WRP)	
		Destination Sequence Distance Vector (DSDV) Routing protocol	
		Fisheye State Routing (FSR)	
		Distance Routing Effect Algorithm for Mobility (DREAM)	Location-based routing.
	Reactive routing	Dynamic Source Routing (DSR) protocol	
		Temporally Ordered Routing Algorithm (TORA)	
		Ad hoc On-demand Distance Vector Routing (AODV) protocol	
		Location Aided Routing (LAR)	Location-based routing.
		Associativity Based Routing (ABR) protocol	Link-stability based routing protocol.
		Signal Stability-base adaptive Routing protocol (SSR)	Link-stability based routing protocol.
Non-uniform routing	Zone-based routing	Zone Routing Protocol (ZRP)	Hybrid routing protocol.
		Hybrid Ad hoc Routing Protocol (HARP)	Hybrid routing protocol also.
		Zone-based Hierarchical Link State routing (ZHLS)	Hybrid routing protocol also.
		Grid Location Service (GLS)	Location service.
	Cluster-based routing	Clusterhead Gateway Switch Routing (CGSR)	
		Hierarchical State Routing (HSR)	
		Cluster Based Routing Protocol (CBRP)	
	Core-node based routing	Landmark Ad hoc Routing (LANMAR)	Proactive routing
		Core-Extraction Distributed Ad hoc Routing (CEDAR)	Reactive routing
		Optimized Link State Routing protocol (OLSR)	Proactive routing

Table 2. Typical multicast routing protocols reviewed in this report

	Tree based	Mesh based	Core based	Group forwarding based
Ad-hoc Multicast Routing (AMRoute)	Y		Y	
Ad hoc Multicast Routing protocol utilizing Increasing id-numberS (AMRIS)	Y			
On-Demand Multicast Routing Protocol (ODMRP)		Y		Y
Core-Assisted Mesh Protocol (CAMP)		Y	Y	
Multicast Ad hoc On-demand Distance Vector (MAODV)	Y			

The protocols are reviewed based on the classification methods previously presented. Additionally, analysis and comparison are given to protocols that belonging to same category. In this way, distinct features, inheriting relationships and performance characteristics of these routing protocols can be easily distinguished and evaluated. In the appendix of this report, the comparison of a couple of unicast outing protocols for mobile ad hoc networks is given. The parameters used for comparison can be grouped into:

- Complexity analysis, including time complexity, communication complexity and storage complexity;
- Mechanism characteristics: such as the route metric, the way for route computation, the way to trigger routing update and the destination for those updates, the need for Hello messages;
- Properties can be used to determine applicable scenarios, fore example, the multiple routes support, the unidirectional link support, the multicast capability, and network structure type.

5.1 Typical proactive routing protocols

5.1.1 The Wireless Routing Protocol (WRP)

The Wireless Routing Protocol (WRP) [32] is a proactive unicast routing protocol for mobile ad hoc networks. WRP uses improved Bellman-Ford Distance Vector routing algorithm. To adapt to the dynamic features of mobile ad hoc networks, some mechanisms are introduced to ensure the reliable exchange of update messages and reduces route loops.

Using WRP, each mobile node maintains a distance table, a routing table, a link-cost table and a Message Retransmission List (MRL). An entry in the routing table contains the distance to a destination node, the predecessor and the successor along the paths to the destination, and a tag to identify its state, i.e., is it a simple path, a loop or invalid. Storing predecessor and successor in the routing table helps to detect routing loops and avoid counting-to-infinity problem, which is the main shortcoming of the original distance vector routing algorithm. A mobile node creates an entry for each neighbor in its link-cost table. The entry contains cost of the link connecting to the neighbor, and the number of timeouts since an error-free message was received from that neighbor.

In WRP, mobile nodes exchange routing tables with their neighbors using update messages. The update messages can be sent either periodically or whenever link state changes happen. The MRL contains information about which neighbor has not acknowledged an update message. If needed, the update message will be retransmitted to the neighbor. Additionally, if there is no change in its routing table since last update, a node is required to send a Hello message to ensure connectivity. On receiving an update message, the node modifies its distance table and looks for better routing paths according to the updated information.

In WRP, a node checks the consistency of its neighbors after detecting any link change. A consistency check helps to eliminate loops and speed up convergence. One shortcoming of WRP is that it needs large memory storage and computing resource to maintain several tables. Moreover, as a proactive routing protocol, it has a limited scalability and is not suitable for large mobile ad hoc networks.

5.1.2 The Destination Sequence Distance Vector (DSDV) routing protocol

The Destination Sequence Distance Vector (DSDV) [4] is a proactive unicast mobile ad hoc network routing protocol. Like WRP, DSDV is also based on the traditional Bellman-Ford algorithm. However, their mechanisms to improve routing performance in mobile ad hoc networks are quite different.

In routing tables of DSDV, an entry stores the next hop towards a destination, the cost metric for the routing path to the destination and a destination sequence number that is created by the destination. Sequence numbers are used in DSDV to distinguish stale routes from fresh ones and avoid formation of route loops.

The route updates of DSDV can be either time-driven or event-driven. Every node periodically transmits updates including its routing information to its immediate neighbors. While a significant change occurs from the last update, a node can transmit its changed routing table in an event-triggered style. Moreover, the DSDV has two ways when sending routing table updates. One is "full dump" update type and the full routing table is included inside the update. A “full dump” update could span many packets. An incremental update contains only those entries that with metric have been changed since the last update is sent. Additionally, the incremental update fits in one packet.

5.1.3 The Fisheye State Routing (FSR)

The Fisheye State Routing (FSR) [33] is a proactive unicast routing protocol based on Link State routing algorithm with effectively reduced overhead to maintain network topology information. As indicated in its name, FSR utilizes a function similar to a fish eye. The eyes of fishes catch the pixels near the focal with high detail, and the detail decreases as the distance from the focal point increases. Similar to fish eyes, FSR maintains the accurate distance and path quality information about the immediate neighboring nodes, and with the progressive detail as the distance increase.

In Link State routing algorithm used for wired networks, link state updates are generated and flooded through the network whenever a node detects a topology change. In FSR, however, nodes exchange link state information only with the neighboring nodes to maintain up-to-date topology information. Link state updates are exchanged periodically in FSR, and each node keeps a full topology map of the network. To reduce the size of link state update messages, the key improvement in FSR is to use different update periods for different entries in the routing table. Link state updates corresponding to the nodes within a smaller scope are propagated with higher frequency.

Using different link state exchange frequencies for nodes with different distances, FSR has better scalability in large networks than traditional Link State routing due to the decreasing traffic overhead. Nevertheless, FSR guarantees the route computation accuracy because when the destination is near, the topology information is more accurate is.

5.1.4 Comparison of WRP, DSDV and FSR

Control traffic overhead and loop-free property are two important issues when applying proactive routing to mobile ad hoc networks. The proactive routing protocols used for wired networks normally have predictable control traffic overhead because topology of wired networks change rarely and most routing updates are periodically propagated. However, periodic routing information updates are not enough for mobile ad hoc routing protocols. The proactive routing in mobile ad hoc networks needs mechanisms that dynamically collect network topology changes and send routing updates in an event-triggered style.

Although belonging to the same routing category for mobile ad hoc networks, WRP, DSDV and FSR have distinct features. Both WRP and DSDV exploited event-triggered updates to maintain up-to-date and consistent routing information for mobile nodes. In contrast to using event-triggered updates, the updates in FSR are exchanged between neighboring nodes and the update the frequency is adjusted according to the nodes' distance effect. In this way, routing update overhead is reduced and the far-reaching effect of Link State routing is restricted.

Different mechanisms are used in WRP, DSDV and FSR for loop-free guarantee. WPR records the predecessor and the successor along a path in its routing table and introduces consistence-checking mechanism. In this way, WRP avoids the forming of temporary route loops with the tradeoff that each node needs to maintain more information and execute more operations. In DSDV, a destination sequence number is introduced to avoid route loops. FSR is a modification of traditional Link State routing and its loop-free property is inherited from Link State routing algorithm.

WRP, DSDV and FSR have the same time and communication complexity. Whereas WRP has a large storage complexity compared to DSDV because more information is required in WRP to guarantee reliable transmission and loop-free paths. Both periodic and triggered updates are utilized in WRP and DSDV; therefore, their performance is tightly related with the network size and node mobility pattern. As a Link State routing protocol, FSR has high storage complexity, but it has potentiality to support multiple-path routing and QoS routing.

5.2 Reactive routing protocols

5.2.1 The Dynamic Source Routing (DSR) Protocol

The Dynamic Source Routing (DSR) [5] is a reactive unicast routing protocol that utilizes source routing algorithm. In source routing algorithm, each data packet contains complete routing information to reach its dissemination. Additionally, in DSR each node uses caching technology to maintain route information that it has learnt.

There are two major phases in DSR, the route discovery phase and the route maintenance phase. When a source node wants to send a packet, it firstly consults its route cache. If the required route is available, the source node includes the routing information inside the data packet before sending it. Otherwise, the source node initiates a route discovery operation by broadcasting route request packets. A route request packet contains addresses of both the source and the destination and a unique number to identify the request. Receiving a route request packet, a node checks its route cache. If the node doesn't have routing information for the requested destination, it appends its own address to the route record field of the route request packet. Then, the request packet is forwarded to its neighbors. To limit the communication overhead of route request packets, a node processes route request packets that both it has not seen before and its address is not presented in the route record field. If the route request packet reaches the destination or an intermediate node has routing information to the destination, a route reply packet is generated. When the route reply packet is generated by the destination, it comprises addresses of nodes that have been traversed by the route request packet. Otherwise, the route reply packet comprises the addresses of nodes the route request packet has traversed concatenated with the route in the intermediate node's route cache.

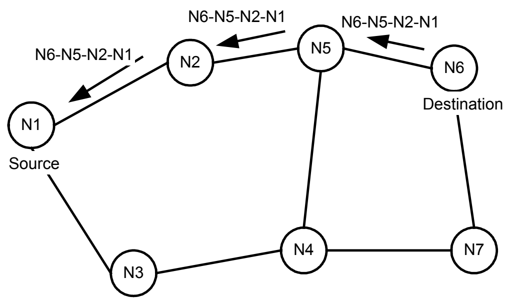

Fig. 1. Route Reply with route record in DSR

After being created, either by the destination or an intermediate node, a route reply packet needs a route back to the source. There are three possibilities to get a backward route. The first one is that the node already has a route to the source. The second possibility is that the network has symmetric (bi-directional) links. The route reply packet is sent using the collected routing information in the route record field, but in a reverse order as shown in Figure 1. In the last case, there exists asymmetric (uni-directional) links and a new route discovery procedure is initiated to the source. The discovered route is piggybacked to the route request packet.

In DSR, when the data link layer detects a link disconnection, a ROUTE_ERROR packet is sent backward to the source. After receiving the ROUTE_ERROR packet, the source node initiates another route discovery operation. Additionally, all routes containing the broken link should be removed from the route caches of the immediate nodes when the ROUTE_ERROR packet is transmitted to the source.

DSR has increased traffic overhead by containing complete routing information into each data packet, which degrades its routing performance.

5.2.2 The Ad Hoc On-demand Distance Vector Routing (AODV) protocol

The Ad Hoc On-demand Distance Vector Routing (AODV) protocol [17] is a reactive unicast routing protocol for mobile ad hoc networks. As a reactive routing protocol, only routing information about the active paths is needed to maintain. In AODV, routing information is maintained in routing tables at nodes. Every mobile node keeps a next-hop routing table, which contains the destinations to which it currently has a route. A routing table entry expires if it has not been used or reactivated for a pre-specified expiration time. Moreover, AODV adopts the destination sequence number technique used by DSDV in an on-demand way.

In AODV, when a source node wants to send packets to the destination but no route is available, it initiates a route discovery operation. In the route discovery operation, the source broadcasts route request (RREQ) packets. A RREQ includes addresses of the source and the destination, the broadcast ID, which is used as its identifier, the last seen sequence number of the destination as well as the source node’s sequence number. Sequence numbers are important to ensure loop-free and up-to-date routes. To reduce the flooding overhead, a node discards RREQs that it has seen before and the expanding ring search algorithm is used in route discovery operation. The RREQ starts with a small TTL (Time-To-Live) value. If the destination is not found, the TTL is increased in following RREQs.

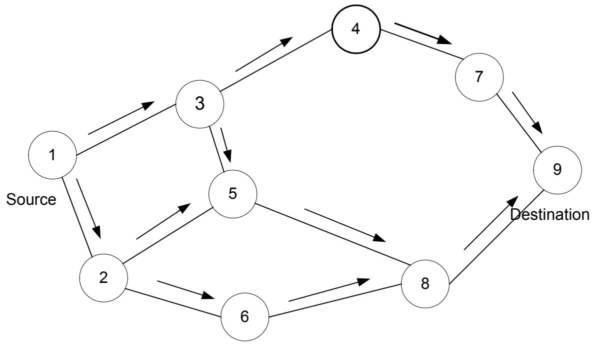

Fig. 2. The Route Request packets flooding in AODV

In AODV, each node maintains a cache to keep track of RREQs it has received. The cache also stores the path back to each RREQ originator. When the destination or a node that has a route to the destination receives the RREQ, it checks the destination sequence numbers it currently knows and the one specified in the RREQ. To guarantee the freshness of the routing information, a route reply (RREP) packet is created and forwarded back to the source only if the destination sequence number is equal to or greater than the one specified in RREQ. AODV uses only symmetric links and a RREP follows the reverse path of the respective RREP. Upon receiving the RREP packet, each intermediate node along the route updates its next-hop table entries with respect to the destination node. The redundant RREP packets or RREP packets with lower destination sequence number will be dropped.

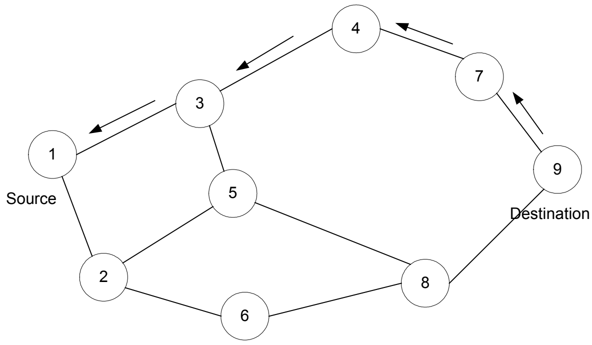

Fig. 3. The forwarding of Route Reply packet in AODV

In AODV, a node uses hello messages to notify its existence to its neighbors. Therefore, the link status to the next hop in an active route can be monitored. When a node discovers a link disconnection, it broadcasts a route error (RERR) packet to its neighbors, which in turn propagates the RERR packet towards nodes whose routes may be affected by the disconnected link. Then, the affected source can re-initiate a route discovery operation if the route is still needed.

5.2.3 The Temporally Ordered Routing Algorithm (TORA)

The Temporally Ordered Routing Algorithm (TORA) [18,19] is a reactive routing algorithm based on the concept of link reversal. TORA improves the partial link reversal method by detecting partitions and stopping non-productive link reversals. TORA can be used for highly dynamic mobile ad hoc networks.

In TORA, the network topology is regarded as a directed graph. A Directional Acyclical Graph (DAG) is accomplished for the network by assigning each node i a height metric hi. A link directional from i to j means $hi > hj$. In TORA, the height of a node is defined as a quintuple, which includes the logical time of a link failure, the unique ID of the node that defines the new reference level, a reflection indicator bit, a propagation ordering parameter and an unique ID of the node. The first three elements collectively represent the reference level. The last two values define an offset with respect to the reference level. Like water flowing, a packet goes from upstream to downstream according the height difference between nodes. DAG provides TORA the capability that many nodes can send packets to a given destination and guarantees that all routes are loop-free.

TORA has three basic operations: route creation, route maintenance and route erasure. A route creation operation starts with setting the height (propagation ordering parameter in the quintuple) of the destination to 0 and heights of all other nodes to NULL (i.e., undefined). The source broadcasts a QRY packet containing the destination's ID. A node with a non-NULL height responds by broadcasting a UPD packet containing the height of its own. On receiving a UPD packet, a node sets its height to one more than that of the UPD generator. A node with higher height is considered as upstream and the node with lower height is considered as downstream. In this way, a directed acyclic graph is constructed from the source to the destination and multiple paths route may exist.

The DAG in TORA may be disconnected because of node mobility. So, route maintenance operation is an important part of TORA. TORA has the unique feature that control messages are localized into a small set of nodes near the occurrence of topology changes. After a node loses its last downstream link, it generates a new reference level and broadcasts the reference to its neighbors. Therefore, links are reversed to reflect the topology change and adapt to the new reference level. The erase operation in TORA floods CLR packets through the network and erase invalid routes.

5.2.4 Comparison of DSR, AODV and TORA

As reactive routing protocols for mobile ad hoc networks, DSR, AODV and TORA are proposed to reduce the control traffic overhead and improve scalability. In the appendix, their main differences are listed.

DSR exploits source routing and routing information caching. A data packet in DSR carries the routing information needed in its route record field. DSR uses flooding in the route discovery phase. AODV adopts the similar route discovery mechanism used in DSR, but stores the next hop routing information in the routing tables at nodes along active routes. Therefore, AODV has less traffic overhead and is more scalable because of the size limitation of route record field in DSR data packets.

Both DSR and TORA support unidirectional links and multiple routing paths, but AODV doesn't. In contrast to DSR and TORA, nodes using AODV periodically exchange hello messages with their neighbors to monitor link disconnections. This incurs extra control traffic overhead. In TORA, utilizing the "link reversal" algorithm, DAG constructs routing paths from multiple sources to one destination and supports multiple routes and multicast [37].

In AODV and DSR, a node notifies the source to re-initiate a new route discovery operation when a routing path disconnection is detected. In TORA, a node re-constructs DAG when it lost all downstream links. Both AODV and DSR use flooding to inform nodes that are affected by a link failure. However, TORA localizes the effect in a set of node near the occurrence of the link failure.

AODV uses sequence numbers to avoid formation of route loops. Because DSR is based on source routing, a loop can be avoided by checking addresses in route record field of data packets. In TORA, each node in an active route has a unique height and packets are forwarded from a node with higher height to a lower one. So, a loop-free property can be guaranteed in TORA. However, TORA has an extra requirement that all nodes must have synchronized clocks. In TORA, oscillations may occur when coordinating nodes currently execute the same operation.

Performances of DSDV, TORA, DSR and AODV are compared in [9] based on simulation. The simulation results showed that DSDV performs well when node mobility rates and speed of movements are low. When the number of source nodes is large, the performance of TORA decreases. As shown in [9], both AODV and DSR perform well for different simulation scenarios. DSR outperforms AODV because it has less routing overhead when nodes have high mobility. A simulation-based comparison of two reactive mobile ad hoc network routing protocols, the AODV and

DSR, is reported in [56]. The general result of [56] was that DSR performs better than AODV when number of nodes is small, lower load and /or mobility, and AODV outperforms DSR in more demanding situations.

5.3 Zone based hierarchical routing protocols

5.3.1 The Zone Routing Protocol (ZRP)

The Zone Routing Protocol (ZRP) [8,20] is a hybrid routing protocol for mobile ad hoc networks. The hybrid protocols are proposed to reduce the control overhead of proactive routing approaches and decrease the latency caused by route search operations in reactive routing approaches.

In ZRP, the network is divided into routing zones according to distances between mobile nodes. Given a hop distance *d* and a node N*,* all nodes within hop distance at most *d* from N belong to the routing zone of N. Peripheral nodes of N are N's neighboring nodes in its routing zone which are exactly *d* hops away from N.

In ZRP, different routing approaches are exploited for inter-zone and intra-zone packets. The proactive routing approach, i.e., the Intra-zone Routing protocol (IARP), is used inside routing zones and the reactive Inter-zone Routing Protocol (IERP) is used between routing zones, respectively. The IARP maintains link state information for nodes within specified distance *d*. Therefore, if the source and destination nodes are in the same routing zone, a route can be available immediately. Most of the existing proactive routing schemes can be used as the IARP for ZRP. The IERP reactively initiates a route discovery when the source node and the destination are residing in different zones. The route discovery in IERP is similar to DSR with the exception that route requests are propagated via peripheral nodes.

5.3.2 The Hybrid Ad hoc Routing Protocol (HARP)

The Hybrid Ad hoc Routing Protocol (HARP) [11] is a hybrid routing scheme, which exploits a two-level zone based hierarchical network structure. Different routing approaches are utilized in two levels, for intra-zone routing and inter-zone routing, respectively.

The Distributed Dynamic Routing (DDR) [11] algorithm is exploited by HARP to provide underlying supports. In DDR, nodes periodically exchange topology messages with their neighbors. A forest is constructed from the network topology by DDR in a distributed way. Each tree of the forest forms a zone. Therefore, the network is divided into a set of non-overlapping dynamic zones.

A mobile node keeps routing information for all other nodes in the same zone. The nodes belonging to different zones but are within the direct transmission range are defined as gateway nodes. Gateway nodes have the responsibility forwarding packets to neighboring zones. In addition to routing information for nodes in the local zone, each node also maintains those of neighboring zones.

As in ZRP, the intra-zone routing of HARP relies on an existing proactive scheme and a reactive scheme is used for inter-zone communication. Depending on whether the forwarding and the destination node are inside the same zone, the respective routing scheme will be applied.

5.3.3 The Zone-based Hierarchical Link State routing (ZHLS)

The Zone-based Hierarchical Link State routing (ZHLS) [10] is a hybrid routing protocol. In ZHLS, mobile nodes are assumed to know their physical locations with assistance from a locating system like GPS. The network is divided into non-overlapping zones based on geographical information.

ZHLS uses a hierarchical addressing scheme that contains zone ID and node ID. A node determines its zone ID according to its location and the pre-defined zone map is well known to all nodes in the network. It is assumed that a virtual link connects two zones if there exists at least one physical link between the zones. A two-level network topology structure is defined in ZHLS, the node level topology and the zone level topology. Respectively, there are two kinds of link state updates, the node level LSP (Link State Packet) and the zone level LSP. A node level LSP contains the node IDs of its neighbors in the same zone and the zone IDs of all other zones. A node periodically broadcast its node level LSP to all other nodes in the same zone. Therefore, through periodic node level LSP exchanges, all nodes in a zone keep identical node level link state information. In ZHLS, gateway nodes broadcast the zone LSP throughout the network whenever a virtual link is broken or created. Consequently, every node knows the current zone level topology of the network.

Before sending packets, a source firstly checks its intra-zone routing table. If the destination is in the same zone as the source, the routing information is already there. Otherwise, the source sends a location request to all other zones through gateway nodes. After a gateway node of the zone, in which the destination node resides, receives the location request, it replies with a location response containing the zone ID of the destination. The zone ID and the node ID of the destination node will be specified in the header of the data packets originated from the source. During the packet forwarding procedure, intermediate nodes except nodes in the destination zone will use inter-zone routing table, and when the packet arrives the destination zone, an intra-zone routing table will be used.

5.3.4 Comparison of ZRP, HARP and ZHLS

As zone based mobile ad hoc network routing protocols, ZRP, HARP and ZHLS use different zone construction methods, which have critical effect on their performance.

In ZRP, the network is divided into overlapping zones according to the topology knowledge for neighboring nodes of each node. In HARP, the network is divided into non-overlapping zones dynamically by DDR through mapping the network topology to a forest. For each node in HARP, the topology knowledge for neighboring nodes is also needed and the zone level stability is used as a QoS parameter to select more stable route. ZHLS assumes that each node has a location system such as GPS and the geographical information is well known, and the network is geographically divided into non-overlapping zones. The performance of a zone based routing protocol is tightly related to the dynamics and size of the network and parameters for zone construction. However, because zones heavily overlap, ZRP in general will incur more overhead than ZHLS and HARP.

All three zone-based routing protocols presented in this subsection use proactive routing for intra-zone communication and reactive routing for inter-zone packet forwarding. Performance of a zone based routing protocol is decided by the performance of respective proactive and reactive routing protocols chosen and how they cooperate each other.

5.4 Cluster-based routing protocols

5.4.1 The Clusterhead Gateway Switch Routing (CGSR)

The Clusterhead Gateway Switch Routing (CGSR) [13] is a hierarchical routing protocol. The cluster structure improves performance of the routing protocol because it provides effective membership and traffic management. Besides routing information collection, update and distribution, cluster construction and clusterhead selection algorithms are important components of cluster based routing protocols.

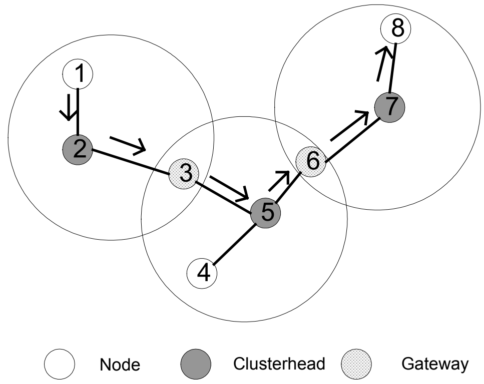

Fig. 4. Cluster structure In CGSR

CGSR uses similar proactive routing mechanism as DSDV. Using CGSR, mobile nodes are aggregated into clusters and a cluster-head is elected for each cluster. Gateway nodes are responsible for communication between two or more clusterheads. Nodes maintain a cluster member table that maps each node to its respective cluster-head. A node broadcasts its cluster member table periodically. After receiving broadcasts from other nodes, a node uses the DSDV algorithm to update its cluster member table. In addition, each node maintains a routing table that determines the next hop to reach other clusters.

In a dynamic network, cluster based schemes suffer from performance degradation due to the frequent elections of a clusterhead. To improve the performance of CGSR, a Least Cluster Change (LCC) algorithm is proposed. Only when changes of network topology cause two clusterheads merging into one or a node being out of the coverage of all current clusters, LCC is initiated to change current state of clusters.

In CGSR, when forwarding a packet, a node firstly checks both its cluster member table and routing table and tries to find the nearest clusterhead along the routing path. As shown in Figure 4, when sending a packet, the source (node 1) transmits the packet to its clusterhead (node 2). From the clusterhead node 2, the packet is sent to the gateway node (node 3) that connecting to this clusterhead and the next clusterhead (node 5) along the route to the destination (node 8). The gateway node (node 6) sends the packet to the next clusterhead (node 7), i.e. the destination cluster-head. The destination clusterhead (node 7) then transmits the packet to the destination (node 8).

5.4.2 The Hierarchical State Routing (HSR)

The Hierarchical State Routing (HSR) [23] is a multi-level cluster-based hierarchical routing protocol. In HSR, mobile nodes are grouped into clusters and a clusterhead is elected for each cluster. The clusterheads of low level clusters again organize themselves into upper level clusters, and so on. Inside a cluster, nodes broadcast their link state information to all others. The clusterhead summarizes link state information of its cluster and sends the information to its neighboring clusterheads via gateway nodes. Nodes in upper level hierarchical clusters flood the network topology information they have obtained to the nodes in the lower level clusters.

In HSR, a hierarchical address is assigned to every node. The hierarchical address reflects the network topology and provides enough information for packet deliveries in the network. Mobile nodes are also partitioned into logical subnetworks corresponding to different user groups. Each node also has a logical address in the form of <subnet, host>. For each subnetwork, there is a location management server (LMS) which records the logical addresses of all nodes in the subnetwork. LMSs advertise their hierarchical addresses to the top level of hierarchical clusters. The routing information, which contains LMSs' hierarchical addresses, is sent down to all LMSs too. If a source node only knows the logical address of the destination node, before sending a packet, the source node firstly checks its LMS and tries to find the hierarchical address of the destination's LMS. Then the source sends the packet to the destination's LMS, and the destination's LMS forwards the packet to the destination. Once the source knows the hierarchical address of the destination, it sends packets directly to the destination without consulting LMSs.

In HSR, logical addresses reflect the group property of mobile nodes and hierarchical addresses reflect their physical locations. Combining these addressing schemes can improve adaptability of the routing algorithm.

5.4.3 Cluster Based Routing Protocol (CBRP)

In the Cluster Based Routing Protocol (CBRP) [12], nodes are divided into clusters and the clustering algorithm is performed when a node joins the network. Before joining, a node is in the "undecided" state. The "undecided" node initiates the joining operation by setting a timer and broadcasts a Hello message. If a clusterhead receives the Hello message, it replies with a triggered Hello message. Receiving the triggered Hello message, the "undecided" node changes its state to "member" state. If the "undecided" node has bi-directional links to some neighbors but does not receive a message from a clusterhead before the local timer generates a timeout, it makes itself a clusterhead. Otherwise, the node remains in "undecided" mode and repeats the joining operation later.

In CBRP, every node maintains a neighbor table in which it stores the information about link states (uni-directional or bi-directional) and the state of its neighbors. In addition to the information of all members in its cluster, a clusterhead keeps information of its neighboring clusters, which includes the clusterheads of neighboring clusters and gateway nodes connecting it to neighboring clusters.

If a source node wants to send a packet but has no active route which can be used, it floods route request to clusterhead of its own and all neighboring clusters. If a clusterhead receives a request it has seen before, it discards the request. Otherwise, the clusterhead checks if the destination of the request is in its cluster. If the destination is in the same cluster, the clusterhead sends the request to the destination, or it floods the request to its neighboring clusterheads. Source routing is used during the route search procedure and only the addresses of clusterheads on the route are recorded. The destination sends a reply including the route information recorded in the request if it successfully receives a route request. If the source doesn't receive a reply in the specified time period, it starts an exponentially backoff algorithm and sends the request later.

The shortening route is proposed in CBRP for performance optimization. Because CBRP uses a source routing scheme, a node gets all information about the route when receiving a packet. To reduce the hop number and adapt to network topology changes, nodes exploit route shortening to choose the most distant neighboring node in a route as next hop.

Another optimization method exploited by CBRP is local repair. Whenever a node has a packet to forward and the next hop is not reachable, it checks the routing information contained in the packet. If the next hop or the hop after next hop in the route is reachable through one of its neighbors, the packet is forwarded through the new route.

5.4.4 Comparison of cluster based hierarchical routing protocols presented

Different clustering algorithms have been introduced to group mobile nodes and elect clusterheads in cluster based routing protocols. In HSR, hierarchical addressing is used and the network may have a recursive multi-level cluster structure. Moreover, a location management mechanism is used in HSR to map the logical address to the physical address. CGSR is based on DSDV, a proactive routing protocol for mobile ad hoc networks, and every node keeps routing information for other nodes in both the cluster member table and the routing table. In CBRP, every node keeps information about its neighbors and a clusterhead maintains information about its members and its neighboring clusterheads. CBRP exploits the source routing scheme and the addresses of clusterheads along a route are recorded in the data packets.

5.5 Core-node based routing protocols

5.5.1 Landmark Ad hoc Routing (LANMAR)

In the Fisheye State Routing protocol (FSR)[1], every node in the network needs to maintain whole network topology information. This strictly limits its scalability. The Landmark Ad hoc Routing (LANMAR) [34] is proposed as a modification of FSR and aims to gain better scalability.

In contrast to FSR, LANMAR belongs to the non-uniform routing category of mobile ad hoc networks. In LANMAR, mobile nodes are divided into predefined logical subnets according to their mobility patterns, i.e., all nodes in a subnet are prone to move as a group. A landmark node is pre-specified for every logic subset to keep track of the subnet.

Using LANMAR, every mobile node has a hierarchical address that includes its subnet identifier. A node maintains the topology information of its neighbors and all landmark nodes, which represent logical subnets. Similar to FSR, neighboring nodes in LANMAR periodically exchange topology information and the distance vector of landmark nodes.

When a source sends packets to the destination inside its neighboring scope (i.e., the source and the destination belong to the same subnet), desired routing information can be found from the source's routing table. Otherwise, the subnet identified in the destination node's address will be searched. Then, according to the distance vector, the packets will be routed towards the landmark node of the logical subset.

Compared to FSR, LANMAR is more efficient because the need to exchange topology information is reduced substantially. However, LANMAR assumes that nodes are grouped into subsets according to their movement patterns and the membership of each subnet remains unchanged during the lifetime of the network, so it is only suitable for specific application scenarios.

5.5.2 The Core-Extraction Distributed Ad Hoc Routing (CEDAR)

The Core-Extraction Distributed Ad Hoc Routing (CEDAR) [22] is a non-uniform routing protocol. In CEDAR, a subset of nodes in the network is identified as the "core". Core is determined according to a distributed algorithm and the number of core nodes is kept to be small. To select core nodes, neighboring nodes periodically exchange link state messages.

Every mobile node in the network must be adjacent to at least one core node and picks this core node as its dominator. The algorithm guarantees that there is a core node at most 3 hops away from another core node. Every core node determines paths to core node nearby using localized broadcasts. The link state information is propagated only among core nodes. The propagation distance of a link state through the network is a function of its stability and bandwidth. Only the state of stable links with high bandwidth is propagated far away and the link state information includes dominators of link endpoints. Hence, in CEDAR, a core node not only knows the state of local links but also the state of stable and high bandwidth links far away.

When a source node wants to send packets to its destination, it informs its dominator core node. Then the dominator of the source finds a route in the core network to the dominator of the destination. This is done by means of a DSR-like route discovery process among core nodes. Then, core nodes involved in the previous step build a route from the source to the destination. Locally available link state information is used according to the QoS requirement such like bandwidth. It is not necessary for the route to include core nodes.

[1] FSR has been presented in subsection 4.1.3.

5.5.3 The Optimized Link State Routing protocol (OLSR)

The Optimized Link State Routing protocol (OLSR) [16] is a proactive non-uniform Link State routing approach. In the original Link State algorithm, each node propagates its link state information to all other nodes in the network. OLSR reduces the overhead and only fewer nodes re-broadcast link state information.

In OLSR, every node transmits its neighbor list using periodical beacons. So, all nodes can know their 2-hop neighbors. Like in CEDAR, OLSR uses an extraction algorithm for multipoint relay (MPR) selection. The multipoint relay set of a node N is the minimal (or near minimal) set of N's one-hop neighbors such that each of N's two-hop neighbors has at least one of N's multipoint relays as its one-hop neighbor. In OLSR, each node selects its MPR independently and only the knowledge of its two-hop neighbors is needed.

When a node broadcasts a message, all of its neighbors will receive the message. Only the MPRs, which have not seen the message before, rebroadcast the message. Therefore, the overhead for message flooding can be greatly reduced.

Using OLSR, each node periodically floods the link state information of its MPR set through the network. The frequency of link state updates is adjusted according to whether a change of the MPR set has been detected. If the MPR set has been changed, the period of link state exchange is set to a minimum value. If the MPR set remains stable, the period is increased until it reaches a refresh interval value. Each node obtains network topology information and constructs its routing table through link state messages. Routes used in OLSR only include multipoint relays as intermediate nodes.

5.5.4 Comparison of CEDAR, OLSR and LANMAR

In a core-node based routing protocol, the core-node extraction method is a key component. CEDAR, OLSR and LANMAR apply totally different approaches for core node extraction purpose.

In LANMAR, the landmark nodes are application related and pre-defined according to their mobility pattern. Obviously, landmark nodes are suitable for tracing groups of nodes that have the same movement patterns. LANMAR is only suitable for specific mobile applications, which meet the assumptions that during the network lifetime, landmark nodes will not change their roles and mobile nodes will not change their mobility patterns.

In CEDAR, a minimal (or nearly minimal) set of core nodes is selected to cover the network according to a certain optimization algorithm. The core nodes can be thought as a dynamically constructed "backbone" as in a cellular network. Core nodes keep link state information of the network. The link state information may include bandwidth, stability or delay information that can be exploited for QoS support and route optimization. The link state propagation is a function of link stability and quality. Only core nodes are involved during route discovery operations. The main disadvantages of CEDAR are that the core extraction algorithm is needed and core nodes have to handle additional traffic associated with route discovery and maintenance.

Different from CEDAR, a node selects its MPR independently in OLSR. A node propagates its MPR set changes through the network, but only MPRs re-broadcast control messages. Thus, OLSR reduces the traffic overhead and improves scalability.

5.6 Routing protocols using location information

5.6.1 Location Aided Routing (LAR)

The Location Aided Routing (LAR) [6] is a reactive unicast routing scheme. LAR exploits position information and is proposed to improve the efficiency of the route discovery procedure by limiting the scope of route request flooding.

In LAR, a source node estimates the current location range of the destination based on information of the last reported location and mobility pattern of the destination. In LAR, an expected zone is defined as a region that is expected to hold the current location of the destination node. During route discovery procedure, the route request flooding is limited to a request zone, which contains the expected zone and location of the sender node.

As shown in Figure 4, there are two different schemes in LAR. In the scheme 1, the source node calculates the expected zone and defines a request zone in request packets, and then initiates a route discovery. Receiving the route request, a node forwards the request if it falls inside the request zone; otherwise it discards the request. When the destination receives the request, it replies with a route reply that contains its current location, time and average speed. The size of a request zone can be adjusted according to the mobility pattern of the destination. When speed of the destination is low, the request zone is small; and when it moves fast, the request zone is large.

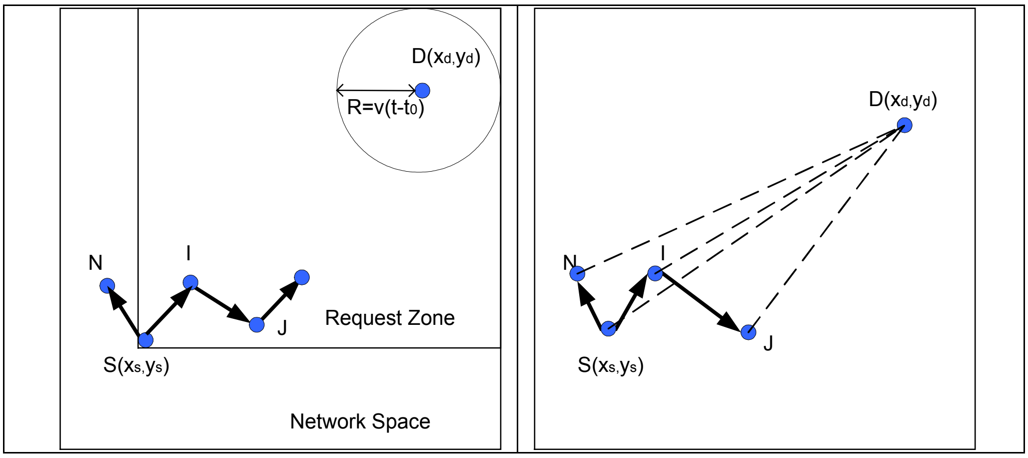

Fig. 5. Two different schemes in LAR

In scheme 2, a source node *S* with coordinate (*xs, ys*) calculates the distance *Dist_s* to the destination *D*, whose coordinate is (*xd, yd*) before it initiates a route discovery operation. Receiving a route request, a node *I* with coordinate *(xi,yi)* calculates its distance *Dist_i* to the destination *D* and forwards the request only if $Dist\_i \leq Dist\_s + \delta$, otherwise it discards the request. Before forwarding the request, node *I* replaces *Dist_s* with *Dist_i*. The non-zero δ increases the success probability of the route discovery procedure.

5.6.2 The Distance Routing Effect Algorithm for Mobility (DREAM)

The Distance Routing Effect Algorithm for Mobility (DREAM) [15] exploits location and speed information of mobile nodes for data packet routing. In DREAM, geographical information is used to limit the flooding of data packets to a small region, rather than to merely provide assistance during the route discovery phase in LAR.

DREAM is a proactive routing scheme. In DREAM, the routing table of a node contains location information of all other nodes in the network. When a source wants to send a packet, firstly it checks its routing table and gets the respective location information of the destination. Then, the source forwards the packet to a neighbor in the direction towards the destination. Therefore, the most substantial issue in DREAM is disseminating the location information through the network. To do that, every mobile node sends location updates comprising its location. The frequency of the location update is determined by the distance and node mobility. Considering the distance effect, nodes departing far away normally have a more stable relative location relationship. Consequently, when a node maintains the location information of another one that is far away, less frequent updates are used. Additionally, each location update is tagged with the "life time" which limits the transmission range of the update. Mobile nodes are allowed to adjust transmission frequencies of their location updates according to their mobility patterns.

5.6.3 The Grid Location Service (GLS)

The Grid Location Service (GLS) [40] provides a distributed location service for routing in mobile ad hoc networks with large number of nodes.

It is assumed in GLS that every node can get its geographic position information by means of GPS or a similar mechanism. Nodes periodically exchange Hello messages with their neighbors. The Hello messages comprise the sender's position and speed information. Hence, every node maintains a table that contains the identifiers and geographic position information of its neighbors. When using GLS with a forwarding scheme, each node also keeps a routing table; the identifier and geographic position of the destination are contained in packet headers.

There are three main operations in GLS: location server selection, location query request and location server update. In GLS, the geographical area is divided into a hierarchy of grids with squares of increasing size. The originate grid cells are called order-1 squares and the next upper level cells are called order-2 squares, and so on. To avoid overlapping, an order-n square is part of only one order-(n+1) square. So, each node is located in exactly one square of each size. A node maintains its geographic position information into a small set of location servers, which are distributed throughout the network. Unique IDs are assigned to nodes in the network and location server set selection is based on node IDs. A node selects the node with the numerically closest ID as its location severs in a square. Location server nodes are not special nodes and each node in the network can act as a location server for some others.

In a routing protocol using GLS to provide location service, before forwarding packets, a mobile node consults its neighbor table and chooses the node closest to the destination. When the next hop node receives the packet, the same operation will continue until the destination is reached. However, a node may have no routing information about the destination when it receives a packet. Such a situation is called a dead-end in location base routing approaches. When a node encounters a dead-end problem, it sends an error message to the source and GLS comes into play.

When a node needs location information of a destination, it initiates a location query request. At first, it searches location servers of the destination in its order-2 square. The request will be sent to the node whose ID is either equal to the ID of the destination or the smallest of the IDs greater than the destination's. This operation will continue if a location server cannot be found in the current level of square order.

When a node moves, it sends location updates to its location servers in the respective squares. A node doesn't need to know the exact locations of its location servers. In location update forwarding, a similar scheme is used as in the location query procedure. The location update frequency is determined according to the distance that a node has moved since last update.

When a node moves into another grid, its location information stored by its location servers becomes obsolete. Additionally, when a location server moves out of a square, the location information becomes useless. To avoid these failures, a node places a forwarding pointer in the grid it is leaving. This forwarding pointer is broadcasted to nodes in the grid and indicates into which grid the departing node is moving.

5.6.4 Comparison of location based routing schemes

Location based routing protocols exploit location and node mobility information for the routing process. LAR, DREAM and GLS use the information in different ways and provide different services.

LAR can be integrated into a reactive routing protocol and its main objective is to perform more efficient route discovery and limit the flooding of route request packets. Using LAR, a sender includes its location in the packets. In contrast to LAR, DREAM itself is a proactive routing protocol and every node keeps location information of all participants in the network. In DREAM, the location update frequency is determined by the relative distance between nodes and their mobility characteristics. GLS is not a routing protocol, but only provides a location service. In GLS, every node has several location servers scattered throughout the network which provide location information.

Although the flooding is constrained in both LAR and DREAM by using location information, they are still not suitable for large-scale ad hoc networks. Their poor scalability roots in the directional flooding reactively initiated in LAR and proactive location information flooding in DREAM. In contrast, GLS can be used in large-scale mobile ad hoc networks with high node density. In GLS, a node chooses a small set of location severs throughout the network. Compared to LAR and DREAM, GLS doesn't exploit flooding for location update and query. Hence, its traffic overhead is greatly reduced. Simulation results in [40, 41] showed that GLS has a high query success ratio in large networks with high node density. However, simulation work in [42] also showed that the performance of GLS greatly declines in small size networks with lower node density.

Because LAR is used for route discovery and GLS provides only location service, they should be used with appropriate location based forwarding schemes [57]. However, DREAM itself is a routing protocol and comprises location service and packet forwarding.

5.7 Link stability based routing protocols

5.7.1 The Associativity Based Routing (ABR) protocol

The overhead of routing protocols for mobile ad hoc networks mainly comes from the dynamically changes of network topology. Some protocols have been proposed aiming to provide routing paths with longevity to reduce the maintenance overhead. The Associativity Based Routing (ABR) protocol [24] uses the degree of association stability as routing metric and tries to find routes that are expected to last longer time.

In ABR, periodic beacons are exchanged between neighboring nodes. Every node keeps an associativity table, in which it records the connection stability between the node and its neighbors over time and space. When receiving a beacon, nodes increase the associativity tick with respect to the sender. Therefore, the link with higher associativity tick is more stable than the one with lower associativity tick. When a neighboring node moves out, the respective associativity tick for it will be reset.

There are three phases in ABR, namely, route discovery, route reconstruction (RRC) and route deletion. When a routing path is needed, the source floods a Broadcast Query (BQ) to initiate the route discovery operation. On receiving a BQ, a node checks if the same message has been seen before. If yes, it discards the BQ. Otherwise, it appends its address and its association ticks inside the BQ packet and broadcasts the BQ. When a subsequent node receives the BQ from its upstream node, it erases the associativity tick entries of its upstream node except the one concerning itself. So, when a BQ arrives at the destination, it contains the addresses of the intermediate nodes and the associativity ticks about all the links along the route. The destination selects the best route according to the longevity of all possible routing paths. The route with the shortest hop number will be chosen if there are multiple paths with the same degree of association stability. Then, the destination sends a reply packet that contains routing information of the selected path back to the source. When an intermediate node forwards the reply packet, it marks its route as valid, therefore, no duplicated data packet will be sent to the destination.

The route reconstruction (RRC) phase may consist of operations like partial route discovery, invalid route erasure, valid route update and new route discovery. If the movement of the source causes a RRC, a new route discovery procedure will be initiated and the source sends a route notification (RN) message to erase the route entries concerning the out-of-date route. When the destination moves, its immediate upstream node erases the routing entry associating with the destination. Then, a localized query (LQ) is sent by the immediate upstream node to test if the destination is still reachable. The hop number from the node to the destination is included in the LQ. If the destination receives the LQ, it selects the best partial route and replies. Otherwise, the next upstream node will repeat the same operation, and an RN message is sent to the next upstream node to erase the invalid route and a LQ is sent to the destination. If the operation has backtracked more than halfway to the source and the destination still cannot be reachable, the source initiates a new BQ process instead of the localized query.

In ABR, a source node initiates the Route Delete (RD) operation by broadcasting a RD message when a route is no longer needed. After receiving the RD message, all nodes along the route delete the respective route entries.

5.7.2 The Signal Stability-Based Adaptive Routing protocol (SSR)

The Signal Stability-based adaptive Routing protocol (SSR) [26] is a reactive routing protocol for mobile ad hoc networks. In SSR, the routing metrics are signal strength and location stability. SSR has two components: the Dynamic Routing Protocol (DRP) and the Static Routing Protocol (SRP). All packets received are processed by DRP to update routing information before being passed to SRP. DRP maintains the Signal Stability Table (SST) and Routing Table (RT).

In SSR, nodes periodically exchange beacons with their neighbors. So, each node knows the signal strength from all its neighboring nodes and stores this information in its SST. When a packet arrives, the SRP checks whether the host node is the destination. If yes, it passes the packet to the high layer. Otherwise, it looks up the RT to find a route to the destination. If there is no route available for the destination, it initiates a route-search operation.

Route-request packets are forwarded only through the strong channels. The destination responds to the first arriving route-request packet with a route-reply message which includes the reversed route to the initiator of route-request packets. The DRPs of the nodes along the path update their RTs accordingly. In this way, the routing paths in SSR have the strongest signal stability because the route-request packets being forwarded over weak channels have been dropped at intermediate nodes.

Sometimes, it is impossible to compose a route merely by strong links and the source will time out before receiving a route-reply. To find a route, the PREF field of route-request header can be set to indicate that links over weak channel are also acceptable.

A route maintenance operation will be initiated whenever an intermediate node detects a link failure. The intermediate node sends an error message informing to the source which channel has failed. Then, the source replies with an erase message to notify all nodes of the broken link. A new route-search procedure is initiated if a route to the destination is still needed.

5.7.3 Comparison of ABR and SSR

In mobile ad hoc networks, node mobility causes link state changes and results in route maintenance operations. Using stability of links instead of hop numbers as metric for routing path selection is a promising solution for reducing control overhead.

Although ABR and SSR are all based on Link State routing algorithm, they have distinct features and different mechanisms. ABR is a reactive routing protocol and is proposed to incorporate the link stability into routing to construct long-lived routing paths. The metric assosiativity is used in ABR to measure how long a wireless link lasts without failure. Following the assumption that the number of the associativity tags of a link reflects how long the link will be available in the future, a route path with greatest associativity tags is constructed. SSR can be seen as an extension of ABR. SSR uses signal stability as routing metric and route requests are propagated only through strong channels. SSR also assumes that the current signal strength of a channel can be used to predict its state in the future. Additionally, in SSR the messages are only propagated through strong channels to reduce the traffic overhead.

5.8 Typical multicast routing protocols

5.8.1 The Ad-hoc Multicast Routing (AMRoute)

The Ad-hoc Multicast Routing (AMRoute) [46] is a tree based multicast routing protocol for mobile ad hoc networks. AMRoute relies on the existence of an underlying unicast routing protocol.

AMRoute has two key phases: mesh creation and tree creation. This protocol can be used for networks in which only a set of nodes supports AMRoute routing function. Using AMRoute, bi-directional unicast tunnels are continuously created between pairs of group numbers that are close together. In contrast to the multicast group members, some nodes for tunnel construction don't support AMRoute. When send a packet to a logically adjacent member, the packet will be physically sent on a unicast tunnel and may pass through many routers. The unicast tunnels form a mesh for each multicast group. AMRoute constructs a multicast distribution tree periodically for each multicast group based on the mesh links available. The group members forward and replicate multicast traffic along the branches of the virtual tree.

In AMRoute, every receiver and sender of a group must explicitly join a multicast group. Each group has at least one logical core that is responsible for member management and tree maintenance. New group members select themselves as cores initially. Each core floods JOIN_REQ messages using expanding ring search to find other group members. When a member of the same group (core or non-core) receives the JOIN-REQ from a core of the same group but a different mesh segment, it replies with a JOIN-ACK and marks that node as a mesh neighbor. A new bidirectional tunnel is established between the core and the responding node of the other mesh. When multiple cores appear after mesh merging, a simple core resolution mechanism is used to select a single core. Members of a group retain knowledge of the current core. If a member receives the TREE_CREATE message from a new core, it resets the current core to the new one only if some deterministic criteria is met (e.g., higher IP address wins). In this way, one mesh is created for each group.

To build a shared tree, each core periodically transmits TREE-CREATE packets to mesh neighbors along the unicast tunnels in the mesh. The period is dependent on the size of the mesh and the node mobility. After receiving a non-duplicate TREE_CREATE message, a group member forwards it on all mesh links except the incoming and marks the incoming and outgoing links as tree links. If a member receives a duplicate TREE_CREATE message, it discards the TREE-CREATE message and sends a TREE-CREATE-NAK message back along the incoming link. The node receiving a TREE-CREATE-NAK marks the link as mesh link instead of tree link. When a node wants to quit the group, it sends JOIN-NAK messages to its neighbors and does not forward any data packets for the group.

Group members detect core failure if no TREE_CREATE message is received from current core within a preset period. After random waiting, the group member designates itself as the new core and generates TREE_CREATE messages and propagates the messages over mesh links.

The robustness comes from the virtual mesh links used to establish the multicast tree in AMRout and a core failure does not prevent data flow. It doesn't need to handle node mobility (done by the unicast protocol) and the non-members do not need to support multicast. AMRoute is efficient by constructing a shared tree for each group. It can operate over any unicast routing protocol; seamlessly over multiple domains and can be extended to wired networks. The major disadvantage of the protocol is that it suffers from temporary loops and creates non-optimal trees when mobility is present.

5.8.2 The Ad hoc Multicast Routing protocol utilizing Increasing id-numberS (AMRIS)

The AMRIS [47] is a proactive shared tree based multicast routing protocol, which is independent of the underlying unicast routing protocol.

The unique feature of AMRIS is that to each node in the multicast session a session specific multicast session member id (msm-id) is assigned. The msm-id provides a heuristic height to a node and the ranking order of msm-id numbers directs the flow of datagrams in the multicast delivery tree. Every node calculates its msm-id during the initialization phase, which is initiated by a special node called Sid. Normally, the Sid is the source node if there is only one source for the session. Otherwise, the Sid is the source node that has the minimum msm-id. The sid broadcasts a NEW_SESSION message to its neighbors. The NEW_SESSION message comprises the Sid's msm-id, the multicast session id, and the routing metrics. After receiving the NEW_SESSION message, a node calculates its own msm-id, which is larger than the one specified in the NEW_SESSION message, but the msm-ids are not consecutive. Before re-broadcast the NEW_SESSION message again, a receiver replace the msm-id field with its own msm-id and the routing metrics of the message. A random jitter is introduced between the reception and re-broadcast of a NEW_SESSION message to prevent broadcast storms.

Every node maintains a Neighbor-Status table. An entry of the Neighbor-Status table stores information for a neighboring node, which includes the unique-id, msm-id, relation (parent/child), remaining timeout value, and routing metric. The Neighbor-Status table is updated based on the contents of NEW-SESSION messages. Additionally, every node is required to broadcast beacons to its neighbors. A beacon message contains the node id, msm-id, membership status, registered parent and child's ids and their msm-ids, and partition id.

When a node joins a multicast session, it firstly determines from received NEW_SESSION and beacon messages which neighboring nodes have smaller msm-ids then itself. Then, it sends a unicast JOIN_REQ to one of its potential parent nodes. If the receiver of the JOIN_REQ already has been a part of the multicast session, it replies with a JOIN_ACK message. Otherwise, it sends a JOIN_REQ_PASSIVE message to its potential parent. If a node fails to receive a JOIN_ACK or receives a JOIN_NAK after sending a JOIN_REQ, it performs a Branch Reconstruction (BR). The BR operation is executed in an expanding ring search until the node succeeds in joining the multicast session.

In AMRIS, the tree maintenance procedure operates continuously and locally to ensure a node's connection to the multicast session delivery tree. If a node has not received any beacon message for a predefined interval of time, it is assumed that a link disconnection has happened. When a link breaks, the node with the larger msm-id (child) is responsible for rejoining. It tries to rejoin the delivery tree by sending JOIN_REQ message to a new potential parent that is one-hop away. If it fails to rejoin the session because that there is no qualified neighbor, the node performs the BR process as described above. In AMRIS, packets loss will happen if a link in the tree breaks until the tree is reconfigured.

5.8.3 The On-Demand Multicast Routing Protocol (ODMRP)

The On-Demand Multicast Routing Protocol (ODMRP) [44, 58] is a reactive mesh based multicast routing protocol. ODMRP uses a forwarding group concept for multicast packet transmission, in which each multicast group G is associated with a forwarding group FG. Nodes in FG are in charge of forwarding multicast packets of group G. In a multicast group of ODMRP, the source manages the group membership, establishes and updates the multicast routes on demand.

Like reactive unicast routing protocols, ODMPR comprises two main phases: the request phase and the reply phase. When a multicast source has a packet to send but it has no routing and group membership information, it floods a Join Request packet to the entire network. Join Request packets are member-advertising packets with piggybacked data payload. When a node receives a non-duplicate JOIN Request, it stores the upstream node ID in its routing table and rebroadcasts the packet. When the

JOIN Request packet reaches a multicast receiver, the receiver refreshes or creates an entry for the source in Member Table and broadcasts JOIN TABLE packets periodically to its neighbors. When a node receives a JOIN TABLE packet, it checks each entry of the table to find out if there is an entry in the table whose next node ID field matches its ID. If there is a match, the node recognizes that it is on the path to the source, thus it is part of the forwarding group. Then it sets the FG_FLAG and broadcasts its own JOIN TABLE built upon matched entries. Consequently, each member of a forwarding group propagates the JOIN TABLE packets until the multicast source is reached via the shortest path. This process constructs (or updates) the routes from sources to receivers and builds a mesh of nodes, the forwarding group.

Multicast senders refresh the member-ship information and update the routes by sending JOIN Request periodically. It uses a soft state approach for group maintenance. Member nodes are refreshed when needed and do not send explicit leave messages.

If nodes in the network have access to geographical information through equipments like GPS, ODMRP then can adapt to node movements by utilizing mobility prediction [58]. With the mobility prediction method, the protocol becomes more resilient to topology changes.

Mobile nodes forward non-duplicated data packets if they are forwarding nodes. Since all forwarding nodes relay data, redundant paths (when they exist) are available for data packets delivery when the primary path is disconnected. ODMRP also operates as an efficient unicast routing protocol, and doesn’t need support from another underlying unicast routing protocol.

5.8.4 The Core-Assisted Mesh protocol (CAMP)

The Core-Assisted Mesh protocol (CAMP) [45, 59] is a proactive multicast routing protocol based on shared meshes. The mesh structure provides at least one path from each source to each receiver in the multicast group.

CAMP relies on an underlying unicast protocol which can provide correct distances to all destinations within finite time. Every node maintains a Routing Table (RT) that is created by the underlying unicast routing protocol. CAMP modifies this table when a multicast group joins or leaves the network. A Multicast Routing Table (MRT) is based on the Routing Table that contains the set of known groups. Moreover, all member nodes maintain a set of caches that contain previously seen data packet information and unacknowledged membership requests.

CAMP classifies nodes in the network as duplex or simplex members, or non-members. Duplex members are full members of the multicast mesh, while simplex members are used to create one-way connections between sender-only nodes and the rest of the multicast mesh. Unlike CBT[60], in which all traffic flows through core nodes, the core nodes in CAMP are used to limit the control traffic when receivers are joining multicast groups.

The creation and maintenance of meshes are main parts of CAMP. A receiver-initiated method is used in the mesh creation procedure. When a node wants to join a multicast mesh, firstly it consults a table to determine whether it has neighbors that are already members of the mesh. If so, the node announces its membership via a CAMP UPDATE. If it does not have such a neighbor, it either propagates a JOIN REQUEST towards one of the multicast group "cores", or attempts to reach a group member by broadcasting requests using an expanding ring search algorithm. Any duplex member of the multicast group can respond to the request with a JOIN ACK, which is propagated back to the request sender.

CAMP exploits special mechanisms (Heartbeat, Push Join) to ensure the validity of all reverse shortest paths (from receiver to source). Periodically, a receiver node reviews its packet cache to determine whether the data packets it has received are from those neighbors that are on the reverse shortest path to the source. If not, the node sends either a HEARTBEAT or a PUSH JOIN message towards the source along the reverse shortest path. This process ensures that all nodes along reverse shortest paths from all receivers to all senders are included in the mesh. Anchors in CAMP are neighbor nodes that are required to rebroadcast any received non-duplicate data packets. Mobile nodes periodically choose and refresh their selected anchors to the multicast mesh by broadcasting updates. A node can leave a group if it is not interested in the multicast session and doesn't act as an anchor for neighboring nodes any more.

5.8.5 The Multicast operation of Ad-hoc On-demand Distance Vector (MAODV)

The Multicast operation of Ad-hoc On-demand Distance Vector (MAODV) is a reactive tree-based multicast routing protocol. MAODV is an extension of the unicast routing protocol Ad-hoc On-demand Distance Vector (AODV).

Using MAODV, all nodes in the network maintain local connectivity by broadcasting "Hello" messages with TTL set to one. Every node maintains three tables, a Routing Table (RT), a Multicast Routing Table (MRT) and a Request Table. RT stores routing information and has the same function as in AODV. Each entry in MRT contains the multicast group IP address, the multicast group leader's IP address, the multicast group sequence number, the hop count to multicast group leader, the hop count to next multicast group member, and the next hops. The next hops field comprises interface and IP address of next hop, the link direction and the activated flag indicating whether the link is added into the multicast tree. Each entry of the Request Table stores the IP addresses of a node, which has sent a request, and the IP address of the requested multicast group. The Request Table provides needed information for optimization.

When a node wants to join a multicast group or sends data packet to a multicast group to which it has no route, it floods a route request (RREQ) message. After receiving an RREP, a node responds to the join request only if it is a member of the desired group. Any node that has fresh route information to the multicast group may respond to the non-join RREQ. A node rebroadcasts the RREQ messages to its neighbors if it is not a member of the desired group or it has no useful routing information for the multicast group. During the broadcasting of RREQ messages, nodes set up pointers to establish the reverse paths. For non-join RREQ messages, the operation is the same as the AODV. For join RREQ, extra operations are needed to maintain the corresponding route entries in MRT. Currently the Enabled flag of this entry is set to FALSE, and it will be set to TRUE only after the node is added to the multicast distribution tree.

After receiving a RREQ message, a member of the desired multicast group responds if the locally recorded group sequence number is fresher than the one contained in the RREQ. The responding node updates its RT and MRT by placing next hop information for the requesting node in the tables. Then, the responding node generates a RREP message and unicasts the message to the source node. The RREP message contains the group sequence number and the IP address of the multicast group leader. Additionally, the Mgroup_hop field in RREP indicates the distance from a source to the multicast group leader. As nodes along the path to the source node receive the RREP, they add entries both in RT and MRT for the node from which they received the RREP. After increasing the hop count and Mgroup_hop fields in the RREP, these nodes continue forwarding RREP towards the source node. Therefore, the forward path is created.

In MAODV, each multicast group has a group leader. A group leader periodically broadcasts "Group Hello" messages through the entire network. The "Group Hello" messages contain IP addresses and corresponding group sequence numbers of the multicast groups of which the sender is the group leader. Nodes update their request table after they received the "Group Hello" messages. Members update their distance to the group leader according to the "Group Hello" messages. MAODV uses a straightforward group leader election method: the first member of a multicast group becomes the leader. This node remains acting as the group leader until it decides to leave the group or a partitioned multicast group merges.

Links in the multicast tree are monitored to detect link breakages, and the downstream node of the break link takes the responsibility for tree maintenance. If the tree cannot be repaired, a new leader for the disconnected downstream nodes is chosen. If a group member initiates the route rebuilding, it becomes the new multicast group leader of the disconnected part. On the other hand, if the initiating node is not a group member and has only one next hop for the tree, it sends its next hop a prune message and leaves the tree. This operation continues until a group member is reached.

When a group member receives a “Group Hello” message for the multicast group and finds that the group leader information contained in the message is different from what it already has, it compares the group leader’s IP addresses. If it is a member of the partition whose group leader has the lower IP address, it initiates reconnection of the multicast tree.

5.8.6 Comparison of multicast routing protocols

For mobile ad hoc networks, the basic idea behind designing multicast routing protocols is to adaptively establish and maintain connections among group members with minimal redundancy and effort. Various algorithms have emerged to achieve this goal using different mechanisms.

Table 3. Comparison of multicast routing protocols for mobile ad hoc netowrks

	AMRoute	*AMRIS*	*ODMRP*	*CAMP*	*MAODV*	*Flooding*
Multicast delivery structure	Tree	Tree	Mesh	Mesh	Core based tree	Mesh
Routing info acquirement /maintenance	Proactive	Proactive	Reactive	Proactive	Reactive	-
Loop free	No	Yes	Yes	Yes	Yes	Yes
Dependency on unicast routing protocol	Yes	No	No	Yes	Yes	No
Control packet flooding	Yes	Yes	Yes	No	Yes	No
Periodic messages requirement	Yes	Yes	Yes	Yes	No	No
Routing hierarchy	Flat	Flat	Flat	Flat	Flat	Flat
Scalability	Fair	Fair	Fair	Good	Fair	bad

Table 3 gives comparison of typical multicast routing protocols for mobile ad hoc networks. Metrics used for comparison are the multicast delivery structure, how to acquire and maintain routing information, whether they are loop-free, the dependency on underlying unicast routing protocol, is the control packet flooding being used, the requirement for periodic control messages, the routing hierarchy and their scalability. As a special case of multicast, flooding is also presented in the table. Additionally, a performance study of various multicast routing protocols can be found in [49].

In mobile ad hoc networks, generally multicast routing protocols based on a mesh structure outperform those based on a tree structure. When node mobility causes link failures, mesh structure provides alternative paths and reduces the loss of data packets. However, mesh based multicast routing protocols normally requires extra overhead for maintenance.

Node mobility may cause formation of route loops, which is critical for mobile ad hoc routing protocols. Except AMRoute, most of multicast protocols discussed previously are loop-free.

Some multicast routing protocols are dependent on certain unicast routing protocols, such as MAODV and CAMP. The former is an extension of the corresponding unicast mechanism, and CAMP relies on the inherent properties of its underlying unicast routing protocol to provide correct and timely distance information.

As in unicast routing protocols for mobile ad hoc networks, multicast routing protocols can be divided into categories as either reactive routing or proactive routing. Among them, AMRoute, AMRIS and CAMP are reactive routing protocols. ODMRP and MAODV are reactive multicast routing protocols. Reactive routing protocols try to reduce overhead by limiting periodically message flooding throughout the network. However, most of existing reactive multicast routing protocols still need periodical network-scale flooding of multicast group membership information.

6. Conclusion

Routing is an essential component of communication protocols in mobile ad hoc networks. This report reviews results in this active research area and presents typical unicast and multicast routing protocols. Additionally, different classification methods are discussed based on different criteria and features to reflect the relationship between existing protocols.

Although fruitful results have been achieved, the survey shows the fact that no single protocol suits all purposes or is widely accepted by a large community. There also many open questions and just nearly investigated areas in research field of mobile ad hoc routing, such as QoS guarantees, adaptability and robustness improvements, multi-path routing and multicast, etc.

New routing schemes are continuously emerging, and at the same time researchers are expending substantial efforts to improve existing routing protocols. It seems that the future success of ubiquitous computing to a large extent will be dependent on the research results in this area.

Appendix:

Table 4 Comparison of unicast routing protocols

	Update destination	Update period	Structure	Route computation	Multicast capability	Hello message requirement	Route metric	Unidirectional link	Multiple routes	Storage complexity	Comm. Complexity	Time complexity
WRP	Neighbors	Hybrid	Flat	Proactive	No	Yes	Shortest path	No	No	O(N*A)	O(N)	O(d)
DSDV	Neighbors	Hybrid	Flat	Proactive	No	No	Shortest path	No	No	O(N)	O(N)	O(d)
FSR	Neighbors	Periodically (different freq.)	Flat	Proactive	No	No	Shortest path	No	May	O(N*A)	O(N)	O(d)
DSR	Source	Event driven	Flat	Reactive	No	No	Shortest path	Yes	Yes	O(d)	O(2N)	O(2d)
TORA	Neighbors	Event driven	Flat	Reactive	No	No	Shortest path	Yes	Yes(DAG)	O(D' *A)	O(2N)	O(2d)
AODV	Source	Event driven	Flat	Reactive	Yes	Yes	Fastest and shortest	No	Yes	O(D')	O(2N)	O(2d)
CGSR	Neighbors and clusterheads	Periodically	Hierarchical	Proactive	No	No	Shortest path	No	No	O(2N)	O(N)	O(d)
ZRP	Neighbors	Periodically	Flat	Hybrid	No	Yes	Shortest path	No	May	O(M+B+D')	O(M)/ O(2B*D)	O(M)/ O(2d)
ABR	Source	Periodically/event driven	Flat	Reactive	No	Yes	Link associativity	No	No	O(d+A)	O(N+y)	O(d+z)
SSR	Neighbors	Periodically/event driven	Flat	Reactive	No	Yes	Signal stability	No	No	O(d+A)	O(N+y)	O(d+z)

d: network diameter; *N: number of nodes in the network;*

y: total number of nodes forming the directed path where the reply packet passes; *A: the average number of neighboring nodes;*

D': the number of maximum desired destinations; *M: the average number of nodes in a zone or cluster;*

B: the average gateway nodes of a zone or clusterhead.

References

[1] IETF Manet charter, http://www.ietf.org/html.charters/manet-charter.html.
[2] Corson S. and Macker J., RFC 2501: Mobile Ad Hoc NETworking (MANET): routing protocol performance issues and evaluation considerations, Internet draft, draft-ietf-manet-issues-01.txt.
[3] S. Ramanathan and M. Steenstrup, A survey of routing techniques for mobile communication networks, Mobile Networks and Applications, 1996, p89-p104.
[4] C. E. Perkins and P. Bhagwat. Highly dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for mobile computers, ACM Computer Communication Review, Vol. 24, No.4, (ACM SIGCOMM'94) Oct. 1994, pp.234-244.
[5] D. Johnson, D. A. Maltz, Dynamic source routing in ad hoc wireless networks, in Mobile Computing (T. Imielinski and H. Korth, eds.), Kluwer Acad. Publ., 1996.
[6] Y. B. Ko and N. H. Vaidya. Location Aid Routing (LAR) in mobile ad hoc networks. In Proc. ACM/IEEE MOBICOM, Oct. 1998.
[7] R. Ramanathan, M. Steenstrup, Hierarchically-orangized, multihop mobile wireless networks for quality-of-service, ACM/Baltzer Mobile Networks and Applications, Vol. 3, No. 1, pp 101-119.
[8] Z. J. Haas. The Zone Routing Protocol (ZRP) for ad hoc networks, Internet Draft, Nov. 1997.
[9] J. Broch, D.A. Maltz, D.B. Johnson, Y.C. Hu, and J. Jetcheva, A performance comparison of multi-hop wireless ad hoc network routing protocols, Proc. MOBICOM, 1998, 85-97.
[10] M. Joa-Ng and I-Tai Lu, A peer-to-peer zone-based two-level link state routing for mobile ad hoc net-works, IEEE on Selected Areas in Communications, vol. 17, no. 8, pp. 1415 1425, 1999.
[11] Navid Nikaein, Christian Bonnet and Neda Nikaein, HARP - Hybrid Ad Hoc Routing Protocol, in proceeding of IST 2001: International Symposium on Telecommunications, Iran/Tehran 2001.
[12] Mingliang Jiang, Jinyang Li and Y. C. Tay. Cluster Based Routing Protocol (CBRP), Internet draft,draft-ietf-manet-cbrp-spec-01.txt.
[13] C. C. Chiang, T. C. Tsai, W. Liu and M. Gerla, Routing in clustered multihop, mobile wireless networks with fading channel, The Next Millennium, The IEEE SICON, 1997.
[14] R.V. Boppana and S.P. Konduru, An adaptive distance vector routing algorithm for Mobile, Ad Hoc Networks, INFOCOM 2001. Proceedings. IEEE, Volume: 3, 2001 Page(s): 1753 - 1762.
[15] S. Basagni, I. Chlamtac, V. Syrotiuk and B. WoodWard, A Distance Routing Effect Algorithm for Mobility (DREAM). Proc. 4th MOBCOM, 1998.
[16] P. Jacquet, P. Muhlethaler, and A. Qayyum, Optimized Link State Routing Protocol, IETF MANET, Internet draft, 1998.
[17] C.E. Perkins and E.M. Royer. Ad hoc on demand Distance Vector routing, mobile computing systems and applications, 1999. Proceedings. WMCSA '99. Second IEEE Workshop on, 1999, p90 - p100.
[18] V. D. Park and M. S. Corson. A highly adaptive distributed routing algorithm for mobile wireless networks, INFOCOM '97, Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Driving the Information Revolution, Proceedings IEEE, Volume: 3, 1997 Page(s): 1405 -1413 vol.3.
[19] V. Park, and S. Corson, Temporally-Ordered Routing Algorithm (TORA) Version 1 Functional Specification. IETF Internet draft, 1997.
[20] Z. J. Haas and M.R Pearlman, The Zone Routing Protocol (ZRP) for ad hoc networks. IETF Internet draft , August 1998.
[21] Hu Zhou and S. Singh, Content Base Multicast (CBM) in ad hoc networks. Workshop on Mobile Ad Hoc Networking and Computing (MobiHoc), August 2000.
[22] P. Sinha, R. Sivakumar and V. Bharghaven, CEDAR: a Core-Extraction Distributed Ad hoc Routing algorithm. IEEE INFOCOM, March 1999.
[23] A. Iwata, C.-C. Chiang, G. Pei, M. Gerla, and T.-W. Chen, Scalable routing strategies for ad hoc wireless networks. IEEE Journal on Selected Areas in Communications, Special Issue on Ad-Hoc Networks, August 1999, p1369-p1379.
[24] C.-K. Toh, Associativity Based Routing For Ad Hoc Mobile Networks . Wireless Personal Communications Journal, Special Issue on Mobile Networking and Computing Systems, p103-p139, March 1997.
[25] Shigang Chen, K. Nahrstedt, Distributed quality-of-service routing in ad hoc networks. IEEE Journal on Selected Areas in Communications, 17(8), August 1999
[26] R. Dube et al., Signal stability based adaptive routing for ad hoc mobile networks, IEEE Pers. Comm., February 1997, p 36-p45.
[27] NS-2 web site, available at http://www.isi.edu/nsnam/ns/
[28] R. A. Meyer, PARSEC User Manual. UCLA Parallel Computing Laboratory, http://pcl.cs.ucla.ed

[29] OPNET Inc. web site, available at http://www.opnet.com/
[30] Ad hoc protocols implementations, http://www.wikipedia.org/wiki/ Ad_hoc_protocols_implementations
[31] Andrew S. Tanenbaum, Computer Networks, 3rd edition, Prentice Hall, 1996
[32] Murthy, S. and J.J. Garcia-Luna-Aceves, An Efficient Routing Protocol for Wireless Networks, ACM Mobile Networks and App. J., Special Issue on Routing in Mobile Communication Networks, Oct. 1996, pp. 183-97.
[33] G. Pei, M. Gerla and T.-W. Chen, Fisheye State Routing in Mobile Ad Hoc Networks. In Proceedings of the 2000 ICDCS Workshops, Taipei, Taiwan, Apr. 2000, pp. D71-D78
[34] G. Pei, M. Gerla, and X. Hong, LANMAR: Landmark routing for large scale wireless ad hoc networks with group mobility. In Proceedings of the ACM Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC), pages 11-18, 2000.
[35] John Jubin and Janet D. Tornow, The DARPA packet radio network protocols. Proceedings of the IEEE, 75(1): 21-32, January 1987.
[36] Gregory S. Lauer, Packet-radio routing. In Routing in Communications Networks, edited by Martha E. SteenStrup, Prentice-Hall, Englewood Cliffs, New Jersey, 1995.
[37] L. Ji and M. S. Corson, A lightweight adaptive multicast algorithm, Proceedings of GLOBECOM'98, November 1998
[38] E. Royer and C.-K. Toh, A review of current routing protocols for ad hoc mobile wireless networks. IEEE Personal Communications, vol. 6, p46-p55, April 1999.
[39] S. Das, C. Perkins, and E. Royer, Performance comparison of two on-demand routing protocols for ad hoc networks. INFOCOM 2000, p3-p12. IEEE.
[40] Jinyna Li, J. Jannotti, D. S. J. De Couto, D. Karger and R. Morris, A scalable location service for geographic ad hoc routing. In Proc. MOBICOM'2000, Boston, MA, USA, 2000.
[41] M. Kaesemann, h. Hartenstein, H. Fuessler and M. Mauve, A simulation study of a location service for position-based routing in mobile ad hoc networks. Technical report, Fakultät für Mathematik und Informatik der Universität Mannheim, Nr. TR-02-007, 2002.
[42] N. K. Guba and T. Camp, Recent work on GLS: a location service for an ad hoc network, Proceedings of the Grace Hopper Celebration (GHC), 2002.
[43] Ching-Chuan Chiang and Mario Gerla, Routing and multicast in multihop, mobile wireless networks. Proceedings of ICUPC '97.
[44] S.J. Lee, M. Gerla, C.C. Chiang, On Demand Multicast Routing Protocol. Proceedings of IEEE WCNC'99, New Orleans, pages 1298-1302, Sept 1999
[45] J.J. Garcia-Luna-Aceves, E.L. Madruga, A Multicast Routing Protocol for Ad-Hoc Networks. Proceedings of IEEE INFOCOM'99, New York, pages 784-792, March 1999.
[46] M. Liu, R. Talpade, A. McAuley, and E. Bommaiah, AMRoute: Ad-hoc multicast routing protocol. Technical Report, CSHCN T. R. 99-1, University of Maryland.
[47] C.W. Wu, Y.C. Tay, AMRIS: a multicast protocol for ad hoc wireless networks. Proceedings IEEE MILCOM'99, Atlantic City, Nov. 1999
[48] E. M. Royer and C. E. Perkins, Multicast operation of the Ad-hoc On-demand Distance Vector routing protocol. Proceedings of ACM/IEEE MOBICOM'99, Seattle, pages 207-218, August 1999
[49] S.J. Lee, W. Su, J. Hsu, M. Gerla and R. Bagrodia, A performance comparison study of ad hoc wireless multicast protocols. INFOCOM 2000, March 2000.
[50] D. Waitzman, C. Partridge and S. Deering. RFC 1075: Distance vector multicast routing protocol. Network Working Group. November 1988.
[51] J. Moy. RFC 1584: Multicast Open Shortest Path First (MOSPF). Network Working Group. March 1994.
[52] B. Fenner, M. Handley, H. Holbrook and I. Kouvelas. Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised). IETF draft. December 2002.
[53] A. Adams, J. Nicholas and W. Siadak. Protocol Independent Multicast - Dense Mode (PIM-DM): Protocol Specification (Revised). IETF draft. February 2003.
[54] J. Moy, RFC 2328: OSPF version 2. Network Working Group. April 1998.
[55] C. Hedrick, RFC 1058: Routing Information Protocol. Network Working Group. June 1988.
[56] S. R. Das, C. E. Perkins, E. M. Royer, Performance comparison of two on-demand routing protocols for ad hoc networks. Proceedings of the IEEE InfoCom, March 2000.
[57]M. Mauve, J. Widmer, H. Hartenstein, A Survey on Position-Based Routing in Mobile Ad-Hoc Networks. IEEE Network, Vol. 15 (2001)
[58] S.J. Lee, W. Su, and M. Gerla, Ad hoc wireless multicast with mobility prediction. In proceedings of IEEE ICCCN99, Boston, MA, OCT. 1999.

[59] J.J. Garcia-Luna-Aceves and E.L. Madruga, The Core-Assisted Mesh Protocol. The IEEE Journal on Selected Area in Communication, vol.17, no.8, Aug. 1999, pp1380-1394.
[60] A. Ballardie, Core Based Trees (CBT version 2) Multicast Routing". Internet Request for Comment 2189, September 1997.
[61] E. M. Royer and C. E. Perkins, Multicast operation of the Ad hoc On Demand Distance Vector Routing protocol. ACM MOBICOM, Aug.1999.