A Distributed Three-hop Routing Protocol to Increase the Capacity of Hybrid Networks

Ze Li and Haiying Shen
Department of Electrical and Computer Engineering
Clemson University
Clemson, SC 29634, USA
{*zxl008, shenh*}*@clemson.edu*

***Abstract*—Hybrid wireless networks combining the advantages of both ad-hoc networks and infrastructure wireless networks have been receiving increasingly attentions because of their ultra-high performance. An efficient data routing protocol is an important component in such networks for high capacity and scalability. However, most routing protocols for the networks simply combine an ad-hoc transmission mode and a cellular transmission mode, which fail to take advantage of the dual-feature architecture. This paper presents a distributed Three-hop Routing (DTR) protocol for hybrid wireless networks. DTR divides a message data stream into segments and transmits the segments in a distributed manner. It makes full spatial reuse of system via high speed ad-hoc interface and alleviate mobile gateway congestion via cellular interface. Furthermore, sending segments to a number of base stations simultaneously increases the throughput, and makes full use of wide-spread base stations. In addition, DTR significantly reduces overhead due to short path length and eliminates route discovery and maintenance overhead. Theoretical analysis and simulation results show the superiority of DTR in comparison with other routing protocols in terms of throughput capacity, scalability and mobility resilience.**

***Keywords*-Distributed routing; Hybrid network**

I. Introduction

A mobile ad-hoc network is a collection of wireless mobile nodes which promises a convenient infrastructure-free communication. In the absence of central control infrastructure, the data is routed to the destination by the intermediate nodes in a multi-hop fashion. Such topology is suitable for a variety of applications such as battle field communication and disaster rescue. However, it is not as reliable as infrastructure networks since the messages are transmitted in the wireless channel and through dynamic routing path.

The infrastructure wireless communication network (e.g. cellular network) is the major communication mode in our daily life. The infrastructure networks excel at inter-cell communication and Internet access. They make possible the support of universal network connectivity and ubiquitous computing by integrating all kinds of wireless devices into the network. In an infrastructure network, nodes communicate with each other through base stations. Because of the long distance one hop transmission feature between base station and mobile nodes, the infrastructure network can provide higher message transmitting reliability and channel access efficiency, but suffers from higher power consumption and single spot failure problem [1]. The growing desire to increase the wireless network capacity for high performance application has produced a significant stimulus to the development of hybrid wireless networks [2], [3], [4], [5], [6], [7]. Hybrid networks can synergistically combine the two types of networks to leverage the advantages of each other in order to increase the throughput capacity of wide-area wireless networks. Wireless devices such as laptop, PDA, smart phones (e.g. Iphone) have both infrastructure interface and ad hoc interface. As the number of these devices increases sharply these years, hybrid transmission structure will be widely used in the near future.

Routing protocol is a critical component that affects the throughput capacity of a wireless network for data transmission. However, few routing protocols have been proposed particularly to meet the needs of a dynamic hybrid network. Most current routing protocols in hybrid networks [2], [7], [6], [8], [9] simply combine the base station transmission mode (i.e. cellular transmission mode) in infrastructure networks and the ad-hoc transmission mode in ad-hoc networks [10], [11]. Specifically, the protocols use multi-hop routing to forward messages to mobile gateway nodes which is closest to the base station or have highest bandwidth to the base station. Functioning to connect the ad-hoc network and infrastructure network, the mobile gateway nodes then forward the messages to the base stations. The direct adoption of the two transmission modes fails to take full advantage of the dual-feature architecture of hybrid networks. In addition, it has a number of problems that are inherently rooted in the ad-hoc transmission mode. The first problem is *high overhead.* Route discovery and maintenance incur high overhead. The wireless random access medium access control (MAC) [12] required in mobile ad-hoc networks, which utilizes control handshaking and back-off mechanism, further increases the overhead. The second problem is *hot-spot generation.* The gateways connecting an ad-hoc network and an infrastructure network could easily become hot spots. In addition, mobile nodes (mobile node and ad-hoc node are interchangeable terms in this paper) only use the channel resources in its route direction. It may generate hot spots while leaving resources in other directions unutilized. Moreover, flooding employed in the mobile ad-hoc routing protocols to discover a new route exacerbates the hot spot problem. Hot spots lead to low transmission rate, severe network congestion, and high data drop rate. The third problem is *unreliability.* Dynamic long routing path leads to unreliable routing. Noise interference and neighbor interference during the multi-hop transmitting process cause a high data dropping rate. Long routing path increases the probability of the occurrence of bath breakdown due to the highly dynamic nature of the ad-hoc networks.

The problems become an obstacle in achieving high capacity and scalability of hybrid networks. Driven by the tremendous advances in wireless networks, there has been an increasingly demand for a scalable routing protocol for hybrid networks to increase their throughout capacity and scalability. This paper presents a Distributed Three-hop Data Routing (DTR) protocol for hybrid networks to increase their capacity and scalability. DTR takes full advantage of the dual-feature architecture of hybrid networks to coordinately integrate the transmission modes in ad-hoc networks and infrastructure networks, and adaptively switch between the two modes based on the QoS requirement of applications. Since most of data traffic from mobile nodes are utilized for inter-cell communication or Internet access based on infrastructures, DTR primarily focuses on inter-cell transmission and Internet access transmissions.

In DTR, a source node divides a message stream into a number of segments (partial packet streams). Each segment is sent to a neighbor mobile node. Based on the QoS requirement, these mobile relay nodes choose between direct transmission or relay transmission to the base station. In the relay transmission, a segment is forwarded to another mobile node with higher capacity to a base station than the current node. In the direct transmission, a segment is directly forwarded to a base station. In the infrastructure, the segments are re-ordered to the original order and sent to its destination. The number of hops in DTR is confined to three, including at most two hops in the ad-hoc transmission mode and one hop in the cellular transmission mode.

Using self-adaptive and distributed routing with high-speed and short-path ad-hoc transmission, DTR significantly increases the throughput capacity and scalability of hybrid networks from three aspects. First, it eliminates overhead caused by route discovery and maintenance especially in a dynamic environment. Second, it alleviates the traffic congestion at the mobile gateway nodes, and meanwhile makes full utilization of channel resources through distributed multi-path relay. Third, it offers high reliability because of the short range transmission. Meanwhile, the implementation of DTR does not bring about much modification on the existing infrastructure architectures

The rest of this paper is organized as follows. Section II presents a review of representative hybrid networks and multi-hop routing protocols. Section III details the DTR protocol, with an emphasis on its routing method, section structure, location management and connection management. Section IV theoretically analyzes the performance of DTR protocol. Section V shows the performance of the DTR protocol in comparison with other routing protocols. Finally, Section VI concludes the paper with remarks on our plans for future work.

II. Related Work

In order to increase the capacity of wireless network, various routing methods for hybrid networks with different features have been proposed [2], [3], [4], [5], [6], [7]. These methods can be primarily classified into two categories. In the first category, the hybrid networks rely on predetermined utilities such as Global Position System (GPS), single direction antennas or fixed proxy relay agent [3], [4], [5] for routing. These predetermined utilities prevent the networks from achieving higher scalability. In the second category, self-organized ad-hoc routing protocols [7], [6], [2], [8], [9] have been integrated with infrastructure network. In these protocols, after building a path to a gateway mobile node or proxy by broadcasting query messages, data messages are forwarded in a multi-hop manner to gateway mobile nodes which forward the messages to the base stations via cellular interfaces. In [8], [9], a node initially communicates with other nodes with ad-hoc transmission mode. The mode switches to the cellular transmission mode when the performance of ad-hoc transmission is worse than the cellular transmission mode. However, these methods are only used to assist intra-cell ad-hoc transmission rather than inter-cell transmission. In inter-cell transmission [7], [6], [2], a message is forwarded via the ad-hoc interface to a gateway mobile node that is closest to or with the highest up-link transmission bandwidth to a base station. The gateway mobile nodes then forward the packets to the base station with the cellular interface. However, most of these routing protocols simply combine routing schemes in ad-hoc networks and infrastructure networks, failing to take advantage of the dual-feature architecture of hybrid networks, and still inherit the drawbacks of the ad-hoc transmission mode.

The ad-hoc routing protocols employed by these routing protocols in the ad-hoc network component includes AODV [10], DSR [11]. The DSR and AODV algorithm determines routes on demand. In DSR, a route is carried in a data packet's header for data transmission. In AODV, each node keeps several routing pathes to the destination after path query process. However, both of these algorithms are prone to path break down and the flooding based path query mechanism lead to high overhead. Using the three hop distributed routing strategy, DTR integrates the two data

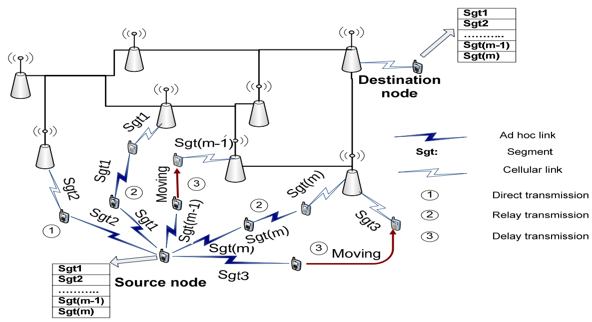

Figure 1. Data transmission in the DTR protocol.

transmission modes coordinately, and meanwhile avoids the drawbacks of the ad-hoc routing protocols.

DTR shares similarity with Two-hop transmission protocol in [13] in terms of the elimination of routing information maintenance in a routing. DTR distinguishes Two-hop in terms of three aspects. First, Two-hop only considers the node transmission within a single cell, while DTR can also deal with inter-cell transmission, which is more challenging and more prevailing than intra-cell communication in the real world. Second, DTR uses distributed transmission involving multiple cells, which makes full use of system resources and avoids bottlenecks. In contrast, Two-hop employs single-path transmission. Third, DTR novelly takes advantage of node mobility for node transmission for QoS based transmission. Grossglauser and Tse [14] proposed a two-hop routing protocol for pure ad-hoc networks with theoretical analysis. In their analytical model, they assumed that every node can tolerate a very long delay and the source node has a huge amount of messages to be sent out, which is not very realistic in practice. DTR is specifically designed for hybrid networks. There are other methods proposed to improve routing in hybrid networks. For example, the work in [15] sets a part of the channel resources particularly for data forwarding. Zadeh and *et al.* [16] proposed to reduce signal attenuation by decreasing power during data transmission. These works are orthogonal to our study in this paper and can be incorporated into DTR to further enhance its performance.

III. Distributed Three-hop Routing Protocol

A. Assumption and Overview

Since base stations are connected with wired backbone, DTR assumes that there is no bandwidth and power constraint in the transmissions between base stations. We use *intermediate nodes* to denote relay nodes that function as gateways connecting an infrastructure network and an ad-hoc network. We assume every mobile node is dual-mode; that is, it has ad-hoc network interface such as WLAN radio interface and infrastructure network interface such as 3G cellular interface.

DTR aims to shift the routing burden on the ad-hoc network to the infrastructure network by taking advantage of widespread base stations in a hybrid network. Figure 1 demonstrates the process of DTR in a hybrid network. We simplify the routings in the infrastructure network for clearness. As shown in the figure, when a source node wants to transmit messages stream to a destination node, it divides the message stream into a number of partial stream called segments and transmits each segment to a neighbor node. Upon receiving a segment from the source node, a neighbor node locally makes choice between direct transmission and relay transmission based on the QoS requirement of applications. These segments are forwarded in a distributed manner to nearby base stations. Relying on infrastructure network routing, the base stations will further transmit the segments to the base station where the destination node resides. Based on the cellular IP transmission method [17], the final base station re-orders the segments into the original order before forwarding the segments to the destination.

The data routing process in DTR can be divided into two steps: uplink from a source node to the final base station, and downlink from the final base station to the data's destination. A critical question in the uplink step is how a source node or relay node chooses nodes for highly efficient segment forwarding, and how to ensure that the final base station sends segments in the right order for a destination node to retrieve the correct original data. Section III-B will present the details for forwarding selection. Section III-C will present the strategy of DTR for the downlink transmission.

B. Uplink Data Routing

DTR limits the path length of the uplink routing to two hops in order to avoid the problems of long-path multi-hop routing in the ad-hoc networks. Meanwhile, DTR always arranges data to be transmitted by high-capacity nodes for high performance routing. In the uplink routing, a source node initially divides its packets streams into a number of segments, then transmits the segments to its neighbor nodes. The neighbor nodes forward segments to base stations, which will forward the segments to the base station where the destination resides.

When choosing neighbors for data forwarding, a source node first chooses the neighbors that have enough space for storing its segment. Then, it chooses neighbors based on the QoS requirement of applications such as nodes' efficiency, reliability and routing speed. For example, delay-tolerant applications (e.g. voice mail, e-mail and text messaging) do not necessarily need fast real-time transmission and may take reliability as priority consideration to ensure successful data transmission. Some applications may take high mobility as priority to avoid hot spots and blank spots. Hot spots are the areas where base station channels are congested, while blank spots are the areas without signals or with very weak signals. In hot spots or blank spots, nodes with high mobility should be the best choices for relay nodes. These relay nodes will quickly move out of the hot spot or blank sport and enter

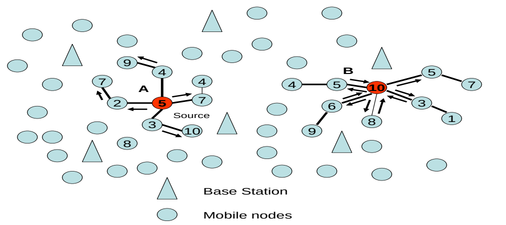

Figure 2. Neighbor Selection in DTR

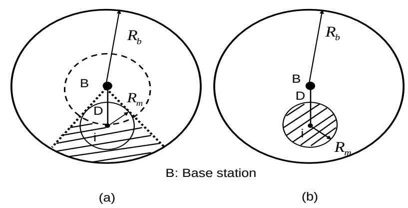

Figure 3. The traffic load in DHybrid and DTR

a cell with high bandwidth of a base station, thus providing efficient data transmission.

Specifically, the source node takes into node capacity for relay node selection. A node's capability includes storage space, bandwidth, CPU, mobility, and etc. For example, if the source node takes reliability as its QoS requirement, the node capacity should be measured by the bandwidth (i.e. channel quality); if the source node takes mobility as its QoS requirement, the node capacity should be measured by the speed of node movement; if the source node takes routing speed as its QoS requirement, the node capacity should be measured by its speed of forwarding data. To measure this node capacity, we use *queue/channel* metric, which is the ratio of a node's message queue size to its channel bandwidth. Smaller *queue/channel* means higher message forwarding speed, and vice versa.

After a neighbor node receives a segment, it uses direct transmission or relay transmission. If the capacity of all its neighbors based on the QoS requirement is no more than itself, the relay node uses direct transmission. Otherwise, it uses relay transmission. In direct transmission, the relay node sends the segment to a base station if it is in the region of a base station. Otherwise, it stores the segment when moving around until it enters the region of a base station. In relay transmission, the relay node needs to choose another relay node based on the QoS requirement. The second relay node will use direct transmission to forward the segment directly to a base station. To choose the second relay node in the relay transmission, the relay node relies on the neighbor selection method described above. As a result, the transmission hops in the ad-hoc network component is confined to no more than two hops. The small number of hops help to increase the capacity of the network and reduce the channel contention in the ad-hoc transmission.

To keep track of the capacity of its neighbors, each node periodically exchanges information with its neighbors about current capacity. In the ad-hoc network, every node periodically needs to send "hello" messages to identify its neighbors. Taking advantage of this policy, nodes piggyback the information along with "hello" messages in order to reduce the overhead caused by the information exchanges. Since a source node's neighbors are more likely to have neighbors with high capacities than the source node, transmitting data segments to neighbors and arranging the neighbors to choose the next hops help to guarantee that data is always transmitted by high-capacity nodes. If a source node has the highest capacity in its region, the segments will be forwarded back to the source node. The source node then forwards the segment to the base stations directly due to the three-hop limit. The latency of this data sending back and forth is negligible because of the high transmission rate of ad hoc interface.

Figure 2 shows an example of neighbor selection in DTR. The value in the node represents its capacity. In scenario A, the source node is in the transmission range of a base station. If the source node directly transmits a message to the base station, the high performance of routing cannot be guaranteed since the source node may have very low capacity. With DTR, the source node sends segments to its neighbors, which further forward the segments to nodes with higher capacities than the source node. In scenario B, the source node has the highest capacity in the area. After receiving segments from the source node, the neighbors forward the segments back to the source node which sends the message to the base station. Thus, DTR always arranges data to be forwarded by nodes with high capacity to the base station. The probability that a segment is lost in DTR is less than traditional transmission methods which do not take into account node capacity in data forwarding.

C. Downlink Data Routing

After a segment is transmitted from an intermediate node to a base station, the base station needs to forward it to another base station which has the location information of the segment's destination. For successful data segment transmission, a base station needs to know the location of the destination node. In a hybrid network, base stations periodically emit beacon signals to locate the mobile nodes. However, destination mobile node switching between different coverage regions of different base stations has posed a challenge for keeping track of the locations of the mobile nodes. For instance, a data is transmitted to base station a which has the information of the data's destination, but the destination has moved to the range of base station b before the data arrives at base station a. To deal with this problem, Cellular IP protocol [17] can be adopted for the node's location tracking. With the protocol, a base station has a Home Agent (HA) and a Foreign Agent (FA). The FA keeps track of mobile nodes moving to other ranges of base stations. HA intercepts in-coming data and then reroutes them to the FA, which then forwards the data to the destination mobile node. The packets are re-assembled at the final base station.

IV. Performance Analysis of the DTR Protocol

In this section, we analyze the effectiveness of the DTR protocol on the enhancement of the capacity and scalability of hybrid networks. The analysis sheds insight into the subtleties of the DTR protocol. We assume that the mobile nodes are distributed uniformly and their positions are independent between each other. The moving directions of these nodes are identical and independently distributed (i.i.d). In *single path transmission*, a message is sequentially transmitted in one routing path. In *multi-path transmission*, a message is divided to a number of segment which are forwarded along multiple pathes in a distributed manner. Since the working frequency of infrastructure networks and ad-hoc networks are different, the communication between a mobile node and a base station through cellular interface will not generate interference to the ad-hoc transmission.

Proposition 4.1: DTR can achieve $O(1)$ throughput per S-D pair. That is, there exists a constant $c > 0$, such that

$$\lim_{n\to\infty} Pr\{\lambda(n) = cp \ \ isfeasible\} = 1, \tag{1}$$

where Pr denotes probability, $\lambda(n)$ is the throughput of the system, and p is the number of S-D paris.

Proof: Suppose in a time slot t, there are $p = O(n)$ pairs of neighbor nodes conducing p concurrent transmissions with negligible inference between pairs. A base station can forward a message to another node immediately with probability 1. Thus, two mobile nodes can be regarded as being virtually connected throughput base stations if they stay in the transmission range of base stations. Therefore, a hybrid network can be regarded as a big virtual ad-hoc network, in which each pair of nodes that stay in the transmission range of base stations are virtually connected with each other. Since n nodes in system are identical and independently distributed, for two arbitrary nodes i and j, the probability that (i, j) can be scheduled for data transmission is $O(1/n)$. In DTR, a message stream is divided into a number of segments which are forwarded by reply nodes to based stations. The throughput of each source-relay pair is $O(1/n)$. Summing up all throughput of $O(n)$ relay nodes for the message stream, the throughput of one S-D pair is $O(1)$. ■

Proposition 4.2: Based on the Point Coordination Function (PCF) mode of the IEEE 802.11 MAC protocol, the lower bound of routing latency of DTR is O$(\frac{T_s}{m})$ and the upper bound is O(T_s), where m is the number of segments of a message, and T_s is the latency of a single-path transmission

Proof: In the PCF mode of the IEEE 802.11 MAC protocol, a time session is divided into a number of time pieces in a round-robin manner. Thus, each time piece for a channel is $t = \frac{T}{N_c}$, where T is the time session and N_c is the number of channels provided by a base station. Therefore, in each time piece, the length of the data that a node can transmit is at most $\frac{T\cdot W}{N_c}$, where W denotes the bandwidth of a channel. In the single-path transmission, the latency for transmitting data with length l is $T_s = \frac{N_c\cdot l}{T\cdot W}$. In the single-path transmission of DTR, the transmission latency is approximately $\frac{N_c\cdot l}{T\cdot W\cdot m}$. Larger bandwidth received by a relay node from a base station leads to shorter latency. Therefore, the upper bound of the transmission latency is $O(\frac{T_s}{m})$. The only situation that the delay of distributed transmission is larger than single-path transmission is when queue/channel ratio is larger than that of the source node. This situation will not happen because DTR guarantees that a segment is forwarded to a node with less than such ratio. Therefore, the worst case of the transmission latency of DTR is that of the single-path transmission, which is $O(T_s)$. ■

Proposition 4.3: In DTR, a source node can find relay nodes for message forwarding with high probability.

Proof: Since the nodes in the system are i.i.d distributed, the probability distribution of the number of mobile nodes that stay in the range of a source nodes with area A_s conforms to a Poission distribution with mean σS, where σ is the average density of nodes in the system. Therefore

$$P(X = k) = e^{-\sigma A_s}\frac{(\sigma A_s)^k}{k!}, \tag{2}$$

where k is the number of mobile nodes staying in the range of a source node. Then, the probability that at least one node is in the area A_s of the source node is $1 - P(X = 0) = 1 - e^{-\sigma A_s}$, which is a monotonically exponentially increasing function. For example, suppose the average number of neighbor nodes of a source node in the area A_s is 6. With the increasing number of mobile devices, such assumption is realistic. Then, the probability of not being able to find any node in the area is $Pr(X = 0) = e^{-6} = 0.25\%$, which is very small. Therefore, in a high-density system, a source node can always find neighbors for message forwarding. ■

DHybrid is used to denote the group of routing protocols in hybrid networks that directly combine the ad-hoc transmission and cellular transmission.

Proposition 4.4: In a hybrid network, the DHybrid routing protocol will lead to load imbalance among the mobile nodes in a cell.

Proof: Figure 3(a) demonstrates a cell, and its shaded region represents all possible positions of the source nodes that choose node i as relay node. The total traffic passing through node i in the system is the sum of the traffic generated by the nodes in A_s. Therefore, the area of shaded region is

$$S = \frac{1}{2}\arcsin\left(\frac{2R_m\sqrt{D^2 - R_m^2}}{D^2}\right)\cdot(R_b^2 - D^2) \quad (0 < D < R_b), \tag{3}$$

where W is the transmission rate of each traffic flow generated by a source node, D is the distance between a base station and relay node i, and R_m and R_b are the transmission ranges of a mobile node and a base station respectively. Therefore, the total traffic passing through node i is

$$W\cdot\sigma\cdot\frac{1}{2}\arcsin\left(\frac{2R_m\sqrt{D^2 - R_m^2}}{D^2}\right)\cdot(R_b^2 - D^2) \quad (0 < D < R_b). \tag{4}$$

Equation 4 shows that the traffic passing through node i decreases as D increases. That is, the nodes close to the base station have more load than the nodes staying at the brim of the cell in the system. ■

Proposition 4.5: In a hybrid network, DTR achieves load balance among mobile nodes in each cell.

Proof: Figure 3(b) shows a cell, and its shaded region represents all possible positions of the source nodes that choose node i as relay node. Since m neighbor nodes are chosen as relay nodes, the traffic from each source node will be normalized to $\frac{W}{m}$. Therefore, the total traffic passing through node i is $\frac{W}{m} \cdot (\sigma \cdot \pi T^2 - 1)$. It shows that the traffic going through relay node i is independent of its location relative to the base station. Since every node in the cell has equal probability to generate traffic, the traffic load is balanced among the nodes in the cell. ■

V. Performance Evaluation

This section demonstrates the properties of DTR through simulations built on ns-2 [18] in comparison with DHybrid [8], Two-hop [13] and AODV [10]. Unless otherwise specified, the simulated network consists of 50 wireless nodes and 4 base stations. In the ad-hoc component of the hybrid network, wireless nodes are randomly deployed around the base stations in a field of 1000×1000 square meters. We use the Distributed Coordination Function (DCF) of IEEE 802.11 as the MAC layer protocol. The transmission range of the cellular interface was set to 250 meters, the transmission power of ad-hoc interface was set to the minimum power required to keep the network connected, and the raw physical link bandwidth was set to 2Mbits/s. We used the two-ray propagation model for the physical layer model. The constant bit rate (CBR) was selected as the traffic mode in the experiment with a rate of 64kbps. In the experiment, we randomly chose some source nodes to continuously send messages to some randomly chosen destination node. The number of channels of each base station was set to 10. We assume that there is no capacity degradation during the transmission in the base stations. This assumption is realistic considering the advanced technologies and hardware in the wired infrastructure networks nowadays. Therefore, we only compare the throughput capacity in the ad-hoc network component of the hybrid network. For each base station, we use the Point Coordination Function (PCF) mode with the IEEE 802.11 MAC protocol.

We employed the random way-point mobility model [14] for each node. In this model, each node moves to a random position with a speed randomly chosen from $(1-20)$m/s. We set the number of segments of a message to the connection degree of the source node.

A. Scalability

Figure 4 shows the average throughput measured by kilobits per second (kbps) per S-D of different routing protocols versus the number of mobile nodes in the system. The Figure shows the throughput of DTR remains almost the same. This result conforms to Propasition 4.1. By fully taking advantage of the spacial reuse with ad-hoc interface, DTR uses distributed multi-path routing to avoid transmission congestion in a single path. DTR avoids the overhead for path query and maintenance, and limits path length to three to avoid the problems of long-path transmission. The throughput of DHybrid decreases as the number of the nodes in the system increases. The reduced throughput of DHybrid is mainly caused by the transmission congestion at the gateway mobile nodes, network partition and neighbor node interference. Since a considerable amount of nodes want to transmit messages to the base station, the nodes close to the base stations which serve as the gateway mobile nodes are easily congested. Meanwhile, the increasing number of mobile nodes in the system leads to high network flow, resulting in frequent route re-transmission. The long transmission path will also leads to high transmission interference which deteriorates the performance of the system.

In Two-hop, because each node also takes full space reuse of the system as DTR, the congestion and signal interference are reduced. Meanwhile, Two-hop enables nodes adaptively switch between direct transmission and relay transmission, gateway nodes are not easily overloaded. Therefore, the throughput of Two-hop is higher than DHybrid. However, since the routing hop of the nodes are confined in one hop, the probability that a better node in its one-hop neighbor will be smaller than DTR. Therefore, the performance of two-hop is worse than DTR, especially in a system with high density of nodes. The reason why AODV has the worst performance is due to its long transmission path.

B. Transmission Delay

Figure 5 shows the transmission delay of different transmission protocols versus network size. Transmission delay is the amount of time it takes for a message to be transmitted from source to destination. From the figure, we can see that DTR generates the smallest delay, which conforms to Proposition 4.2. The delay of DHybrid is $5-6$ times larger than DTR, which is approximately the connection degree of the nodes in the system. As the number of nodes in the system increases, the connection degree of each node is increased, and the ratio of delay time of DHybrid to DTR increases. DTR also produces shorter transmission delay than Two-hop. This is caused by two reasons. First, the multi-path parallel routing of DTR saves much transmission time as the proof of Proposition 4.2 explained. Second, the distributed routing of DTR enables some messages to be forwarded to the neighbor cell with good transmission channels rather than waiting in the current hot cell for a transmission channel. We can also observe that the performance of Two-hop is better than DHybrid. It is because the multi-hop transmission component of DHybrid results in a

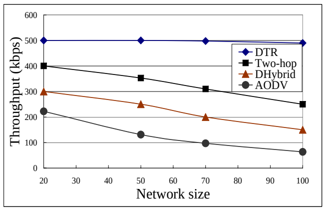

Figure 4. Throughput (simulation).

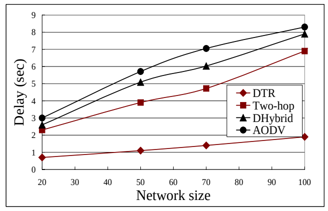

Figure 5. Transmission delay.

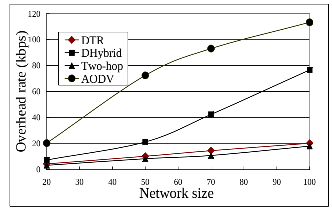

Figure 6. Overhead versus network size

higher delay due to the queue delay in each hop. Because of the long distance transmission without the support of an infrastructure network, AODV generates the longest delay.

C. Communication Overhead

We use the generation rate of control messages in the network and MAC layer in kbps to represent the communication overhead of the system. Figure 6 illustrates the communication overhead of DTR, Two-hop, DHybrid and AODV. We can see that the communication overhead between DTR and Two-hop is very close. It is because both DTR and Two-hop are transmission protocols of short distance and small hops. The reason why DTR has slightly higher communication overhead than Two-hop is because the DTR utilizes three hop transmission which have one more hop than two hop transmission. However, the marginal overhead increase leads to a much higher transmission throughput as shown in Figure 4. DHybrid has much larger overhead than DTR and Two-hop is because of its high overhead of routing path maintenance. The pure AODV routing protocol results in much more overheads than others. Without a infrastructure network, the packets in AODV travel a long way from the source node to the destination node through a path much longer than DHybrid's.

D. Effect of Mobility

In order to see how the node mobility influences the performance of the routing protocols, we evaluated the throughput of these four transmission protocols with different node mobility. Figure 7 plots the throughput of DTR, DHybrid, Two-hop and AODV versus node moving speed. From the figure, we can see that the increasing mobility of the nodes does not adversely affect the performance of DTR and Two-hop. It is intriguing to find that high mobility can even help DTR to increase its throughput. It is because DTR and Two-hop transmission mode do not need to maintain their routes and routing tables, thus the network partition and topology changes of the system will not affect the transmission of DTR and the Two-hop. Moreover, as the mobility increases, a mobile node can meet more other nodes in a short time period. Therefore, DTR enables the segments to be quickly sent to a high-capacity nodes or to the nodes in a less-congested cell. As node mobility increases, the throughput of DHybrid decrease. In DHybrid, the messages are routed in a multi-hop fashion. When the links between nodes are broken because of node mobility, the messages are dropped. Therefore, when nodes have smaller mobility, the link between the mobile nodes last longer. Thus, more messages can be transmitted. Hence, the throughput of DHybrid is affected by node mobility. However, since the DHybrid can adaptively adjust their routing between ad-hoc transmission and cellular transmission, the performance of DHybrid is much better than AODV. With no infrastructure network, AODV leads to much lower throughput than others. Its throughput also drops as node mobility increases due to the same reasons as DHybrid.

E. Load Distribution

In this experiment, we test the load distribution of DTR, DHybrid and Two-hop in the hybrid network environment. We define the load rate as the distribution of the number of packets sent to the base station. We normalize the distance from a mobile to its base station according to the function $\frac{d}{R}$, where d is the actual distance and R is the radius of its cell.

The space of the cell are divided into several concentric-circles. The loads of the nodes in each circle are measured to show the load distribution. Figure 8 shows the average load of a node corresponding to the normalized distance from itself to its base station. The figure shows that most of the traffic load of DHybrid are located at nodes near the base station. The nodes far from the base station have very low load. The results conform to Proposition 4.4. In DHybrid, if a source node wants to access the Internet backbone or engage in inter-cell communication, it transmits the messages to the base stations in a multi-hop fashion. Therefore, the nodes near the base stations will have most load. On the other hand, since there is little traffic going through the nodes at the brim of a cell, the load of these nodes are small. As a result, some nodes can easily become hot spots while the resources of other nodes are not fully utilized. The load imbalance prevents DHybrid from fully utilizing the system resources. The traffic load of DTR is almost evenly distributed in the system which is in line with Proposition 4.5. In DTR, the traffic from the source node are distributed among a number of relay neighbors for further data forwarding. The nodes at the brim of the cell also take responsibility for the messages forwarding since the neighbor nodes of these brim node could be located in other cells with good transmission channel. In Two-hop,

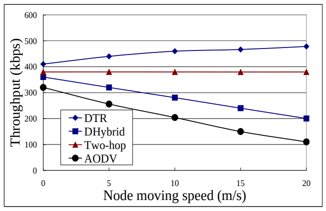

Figure 7. Throughput versus speed.

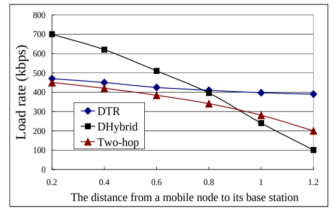

Figure 8. Load distribution

the source node considers direct transmission or one-hop relay transmission based on the channel condition. Since the node is chosen within one hop, the messages will not gather close to the base station because of the limited transmission range. However, because of the sequential transmission, the messages is not likely to be transmitted out of the current cell. Therefore, the two hop protocol cannot reach a good load balance.

VI. Conclusions

Current hybrid networks simply combine the routing protocols in the two types of networks for data transmission, which prevents them from achieving higher system capacity. In the paper, we propose a Distributed Three-hop Routing (DTR) data routing protocol that synergistically integrates the dual features of hybrid networks in the data transmission process. In DTR, a source node divides a message stream into segments and transmits them to its mobile neighbors, which further forward the segments to their destination through infrastructure network. DTR limits the routing path length to three, and always arranges high-capacity nodes to forward data. Unlike most existing routing protocols, without route discovery and maintenance, DTR renders significantly lower overhead. In addition, its distinguished characteristics of short path length, short-distance transmission and widespread load distribution provide high routing reliability and efficiency. Theoretical analysis and simulation results show that DTR can dramatically improve the throughput capacity and the scalability of hybrid networks due to its highly scalable, efficient, reliable and low-overhead features.

Acknowledgment

The authors are grateful to the anonymous reviewers for their valuable comments and suggestions. This research was supported in part by U.S. NSF grants CNS-0834592 and CNS-0832109.

References

[1] R. S. Chang, W. Y. Chen, and Y. F. Wen. Hybrid wireless network protocols. *IEEE Transaction on Vehicular Technology*, 2003.

[2] H Luo and et al. Ucan: A unified cell and ad-hoc network architecture. In *Proc. of MOBICOM*, 2003.

[3] P. K. McKinley, H. Xu, and et al. Unicast-based multicast communication in wormhole-routed direct networks. *TPDS*, 5:252–265, 1992.

[4] H. Wu, C. Qiao, S. De, and O. Tonguz. Integrated cell and ad hoc relaying systems: iCAR. *J-SAC*, pages 2105– 2115, 2001.

[5] Y. H. Tam, H. S. Hassanein, S. G. Akl, and R. Benkoczi. Optimal multi-hop cellular architecture for wireless communications. In *Proc. of LCN*, 2006.

[6] Y. D. Lin and Y. C. Hsu. Multi-hop cellular: A new architecture for wireless communications. In *Proc. of INFOCOM*, 2000.

[7] P. T. Oliver, Dousse, and M. Hasler. Connectivity in ad hoc and hybrid networks. In *Proc. of IEEE Inforcom*, 2002.

[8] B. Liu, Z. Liu, and D. Towsley. On the capacity of hybrid wireless networks. 2003.

[9] H. Y. Hsieh and R. Sivakumar. A hybrid network model for wireless packet data networks. In *Proc. of GLOBECOM*, 2002.

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

[11] D. B. Johnson and D. A. Maltz. Dynamic source routing in ad hoc wireless networks. *IEEE Mobile Computing*, 1996.

[12] S. Xu and T. Saadawi. Does the ieee 802.11 mac protocol work well in multihop wireless ad hoc network? *Communications Magazine, IEEE*, 2001.

[13] Y. Wei and D. Gitlin. Two-hop-relay architecture for next-generation WWAN/WLAN integration. *IEEE Wireless Communication*, 2004.

[14] M. Grossglauser and D. Tse. Mobility increases the capacity of ad hoc wireless networks. In *Proc. of TON*, 2002.

[15] X. Wu, S.G. Chan, and B. Mukherjee. Madf: A novel approach to add an ad hoc overlay on a fixed cell infrastructure. In *Proc. of WCNC*, 2000.

[16] A. N. Zadeh and et al. Self- organizing packet radio ad hoc networks with overlay (SOPRANO). *IEEE Communications Magazine*, pages 149–157, 2002.

[17] A. G. Valko. Cellular ip: A new approach to internet host mobility. *ACM Computer Comminication*, 1999.

[18] The network simulator - ns-2.