Carleton University, Systems and Computer Engineering, Technical Report SCE-04-14, August 2004

Survey on Mobile Ad Hoc Network Routing Protocols and Cross-Layer Design

Liang Qin and Thomas Kunz
Systems and Computer Engineering
Carleton University
August 2004

A "mobile ad hoc network" (MANET) is an autonomous system of mobile routers connected by wireless links. The routers are free to move randomly and organize themselves arbitrarily; thus, the network's wireless topology may change rapidly and unpredictably. Such a network may operate in a standalone fashion, or may be connected to the Internet. Multi hop, mobility, large network size combined with device heterogeneity, bandwidth and battery power constrain make the design of adequate routing protocols a major challenge. In recent years, many routing protocols have been proposed for MANET. Basically these protocols can be fit in one of two major categories: on-demand such as AODV [1] and DSR [2], and proactive such as DSDV [3] and OLSR [4]. The review and performance comparison of these protocols are in [5][6][7]. A more comprehensive survey can be found in [8]. In this survey, we will not focus on individual routing protocols; instead we will discuss some new ideas proposed recently mainly to improve MANET throughput and scalability in different ways with some new routing metrics, new technologies such as multi-rate, multi-channel and hierarchical structure, by using cross-layer design.

1. Cross-layer Design of Multi-hop Wireless Networks

In most networking software, the protocols are divided into several modules to form a protocol stack. Each layer makes use of the services provided by the layer directly below it, and also provides service to the layer directly above it. The communication is limited between adjacent layers with a minimum set of primitives. The layering principle simplifies design and implementation and provides the possibility of alternative layer implementations. The success of the Internet demonstrates the power of a layered design in wired networks. The characteristics of wireless networks differ from wired networks in several ways, caused by their low link capacity and high bit error rates:

- Due to small-scale and large-scale channel variation, the channel quality changes within milliseconds depending on the node's location and mobility. The routing protocol cannot select a route simply based on a single route request message [9].
- The wireless link capacity depends on the status of other links in its transmission range. Therefore the congestion can also be caused by the inference of other links.

Because of the direct coupling between the physical layer and the upper layers, the traditional protocol stack is not sufficient for wireless networks. Cross-layer design methodology is an active research area to improve wireless network performance, where the information is exchanged between different protocol layers dynamically. In a wireless network, physical layer, MAC layer and routing layer together contend for the network resource. The physical layer affects MAC and routing decisions by its transmission power and rate. The MAC layer is responsible for scheduling and allocating the wireless channel, which finally will determine the available bandwidth of the transmitter and the packet delay. This bandwidth and packet delay also can affect the decision at the routing layer to select the link. The routing layer chooses the wireless links to relay the packets to the destination. The routing decision will change the contention level at the MAC layer, and accordingly the physical layer parameters.

There are some existing examples for cross-layer design in wireless networks. In CDMA2000 HDR (high data rate)[10], each node periodically measures the quality of the channel to the base station and sends it to the base station, so the base station could give priority to the users with better channel quality to improve the throughput. In [11] the cross-layer design addresses the joint problem of power control and scheduling for multi-hop wireless networks with QoS. It takes SINR and minimum rate as constraints to minimize the total transmit over the links. [12] studied the interaction of the routing protocols and MAC protocols for wireless ad hoc networks under different mobility parameters. Experiments have been run with difference parameter combinations of routing protocols (AODV, DSR and LAR Scheme 1 [13]), MAC layer (MACA, IEEE 802.11 and CSMA), speed of nodes and data packet injection rates. A statistical technique, ANOVA (analysis of variance), is used to analyze the results. The results show significant interacts between these variables in terms of performance [12].

Table 1. Statistical Results on Interaction between Various Input Variables

Performance Metric	Interaction	Input Variables
Latency	3-way	Routing protocols, node speeds and MAC
Packet received	4-way	Routing protocols, node speed, Injection speed and MAC
Long term fairness	2-way	Routing/MAC protocol, MAC/Injection rate

[14] is a simple cross-layer design example. AODV routing protocol is used in an ad hoc network for transmitting real-time video. The routing information created by AODV can be shared with application programs. When a sender wants to send packets, it will check the information first. If the route changed in terms of hop counts, it will adjust encoding bit rate to adapt to the links condition. In [15] the congestion information of mobile node is used by different layers such as network, transport and higher layer. A mobile node can be measured its congestion level by two metrics: one is the transmission queue length; the other is MACV layer utilization level, which can be obtained by monitoring the busy level of the wireless medium around it. Within network layer, the congestion parameter can be used as a routing metric to select route for the proactive routing protocols, or node can change the time interval to advertise its routing information; for reactive routing protocol such as DSR, when a node may not rebroadcast Route Request messages if it knows the medium around it is busy. Within the transport layer, a node may set Explicit Congestion Notification (ECN) bits in a packet's IP header. Within other higher layer protocols, for example, the sender may compress the data before transmission if it knows the some link in the route is very busy. When applying theses uses in DSR, the simulation results show substantial improvement in terms of delivery radio, overhead and scalability.

While cross-layer design may gain some improvement on network performance, it also may lead to negative consequences. When we break the layer isolation in the protocol stack, we also loose the design abstraction. [16] discusses the potential risk of cross-layer design. The authors state that though we can gain some improvement on network performance, this gain is not unbounded.

Any design changes in the protocol stack when adding interaction between different layers may have effect on the whole system, which may lead to "spaghetti" design. Though we may have short-term gain, because of unforeseen reactions in the system, further changes become more difficult. The authors provide some design principles [16]:

A. Interactions and The Law of Unintended Consequences: when we add interaction to different layers, we must consider the effect on other layers in the systems.

B. Dependency Graph: representing the interaction between protocol parameters as graph.

C. Time-Scale Separation and Stability: From dependency graph, we can derive some stability principles.

D. The Chaos of Unbridled Cross Layer Design: more design issues need to be considered, for example, the code maintenance.

The authors also give some examples to illustrate the potential problems in cross-layer design. We will discuss one example in Section 3.

2. Problems with the IEEE 802.11 MAC Protocol

The IEEE 802.11 medium access control (MAC) is a standard widely used in wireless LAN and wireless ad hoc networks. However it was not designed for multi-hop wireless ad hoc networks, and there are some papers on this topic. In [17], the authors set up a simple wireless network with 8 static nodes in a string topology. In the network, each node is 200m apart, and the wireless interface's transmission range is 250m and TCP traffic is used. After simulations in NS2, the authors found the following three problems rooted in the MAC layer:

A. The TCP instability problem: the throughput of only one TCP connection existing in the network repeatedly reached or was near zero. It was caused by the interactions between different nodes carrying TCP-data and TCP-ACK traffic. The "hidden node problem" causes collision in an intermediate node, and the "exposed node problem" prevents the intermediate node from sending a CTS message. So the node cannot reach its neighbor and the link is often broken in the middle of the route. Using smaller maximum window size can lessen or clear this problem.

B. Serious unfairness: one kind of unfairness is called "neighboring node one-hop unfairness". Caused by the same problems as A, when there are two TCP connections in the network, one session may be completely shut down and have no chance to restart in some circumstances even if it starts much earlier. This problem cannot be solved by adjusting the window size.

C. Incompatibility problem: two TCP sessions cannot coexist in the network at the same time, and the turnover time is totally random, which is caused by the "exposed node problem". It cannot be solved by adjusting TCP parameters.

In the wireless LAN with infrastructure, using RTS/CTS and carrier sense to prevent "the hidden node problem" seems work well, because the nodes that potentially interfere with the reception at the receiver are all in the receiver's sensing range under the assumption that the transmission range is the same as the sensing range. It is not true for multi-hop ad hoc networks. In the simulation tools which model WaveLAN cards, the sensing range is twice longer than the successful transmission range, so that some nodes cannot send back CTS packets when their neighbors are transmitting, which leads to broken links. [18] proposes an adaptive RTS/CTS mechanism to reduce the unfairness caused by IEEE 802.11. Because we cannot totally turn off RTS/CTS, it will still result in unfairness and collision. In an adaptive RTS/CTS scheme, a node will turn off RTS/CTS when the number of "Waiting for CTS timeout" events exceeds a threshold. The counting number is updated in a sliding window fashion. The simulation results show that this adaptive mechanism can significantly improve the fairness both for UDP and TCP transmission. [19] explores the RTS/CTS issue even further. This paper discovers there are some scenarios where RTS/CTS can induce congestion in ad hoc networks. At first CTS/RTS may cause a blocking problem, as illustrated in Figure 1 [19]. Node B is sending packets to node A. Node C receives both RTS and CTS, so it will stop transmitting. If at this time node D sends RTS to node C, node C cannot reply with CTS, finally node D will enter into exponential backoff mode. In this scenario, node C need not be either a hidden node or an exposed node as Figure 1 shows, because it can receive both RTS and CTS. In the current implementation of the RTS/CTS mechanism, when a node received an RTS packet not addressed to it, it is required to stop transmitting. In the blocking problem scenario, these nodes neighboring to the blocked node may be falsely blocked, and even worse, the false blocking may spread through the network until some event like packet drop breaks this kind of pseudo-deadlock. [19] proposes a solution to the false blocking problem. The basic idea is RTS validation: when a node hears RTS which is not addressed to itself, it will defer a certain amount of time to check if there are really data packets in transmission. If the medium is still idle, which means that false blocking may happen, it will not defer any more. The simulation results show this solution can significantly improve the throughput.

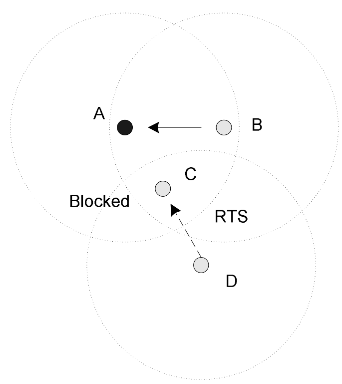

Figure 1. Blocking Problem: node C is blocked due to the communication between node A and node B. Therefore, node D does not get any response to the RTS packets it sends and enters backoff.

[20] discusses a performance anomaly of IEEE 802.11 caused by CSMA/CA, which provides equal probability for each node to access the channel in long term. In other words, every node has the same chance to send a packet at the speed based on its data rate. If there are nodes with different data rates in the same cell, the throughput of all nodes with higher data rates will be reduced to the level of the lower rates. The paper also found that when using TCP, the packet loss of the node which causes degraded bit rates will reduce its sending rate, which in turn gives more capacity to other nodes; finally the negative effect is alleviated.

Another characteristic of IEEE 802.11, which has an effect on ad hoc wireless routing protocols using broadcast messages to sense neighboring nodes, are the so called "communication gray zones" [21]. Here we take AODV as an example to explain this phenomenon. AODV is a reactive protocol. It uses broadcast messages to discover routes and periodically broadcasts HELLO beacons to detect neighboring nodes so that it can update routes in the routing table. It uses unicast to send data packets and route reply packets. The HELLO messages have special properties contributing to the “gray zone” creation [21]:

A. Different Transmission Rate: In IEEE 802.11a/b/g, broadcast packets are always transmitted at a base bit rate while data packets can be sent at higher rate. Then the broadcasted packets can reach further than data packets. This is the main reason causing a “gray zone”.
B. No Acknowledgements: Broadcast messages do not require ACKs. This means that received HELLO messages do not indicate it is a bidirectional link.
C. Small Packet Size: for a weak link, a HELLO message has higher successful transmission rate than bigger sized data packets.
D. Fluctuating Links: At the edge of the transmission range, the link quality is poor and unsteady. If the HELLO messages are received successfully, this link becomes an unreliable link in a route.

The network will experience severe packets loss when some link is inside a gray zone. This occurrence has not been observed in NS2 simulation before, because in NS2 all the packets (unicast or broadcast) are transmitted at 2Mbit/s rate, and treat the link as an on/off switch, eliminating the link fluctuation property actually existing in the real world.

In [21], the authors provide three modifications to AODV-UU (a new implementation of AODV): exchanging neighbor sets, N-consecutive HELLOs, and SNR threshold for control packets. The simulation results show that applying SNR threshold method on AODV, which raises the receiving threshold for control messages, can reduce more packet loss than the other two methods; on the other hand this threshold is context specific and sometimes may cause the loss of some acceptable links. The simulation results of AODV, OLSR [4], and LUNAR [22] are shown in Table 2 [21].

Table 2. Comparison against OLSR and LUNAR for all three experiments ("Roaming node" scenario)

Protocol	success ratio		HTTP cycles
	Ping	MP3	
OLSR	89.0%	91.9%	32.5
LUNAR	96.5%	96.8%	31.5
AODV-UU	91.9%	97.9%	33
AODV-UU+SNR	99.1%	99.7%	34

3. Shortest Path is not Enough: Problems in Ad Hoc Network Routing Protocols

Existing wireless ad hoc routing protocols commonly use minimum hop counts as metric to find routes, and under two assumptions: a link which is good for route discovery messages is still good for data packets; secondly, the link quality is binary: either very good or very bad. So protocols such as DSR and AODV use broadcast messages to find the shortest paths, when the node receives the route reply, it will use this route to transmit data. [23] uses experimental evidence from two wireless test beds to show that using minimum hop counts as metric often leads to less capacity than the existing best paths, because:

- The link quality is spread out.
- Some links are asymmetric
- Link quality varies over time.

- There is no good correlation between link signal strength and delivery rate.

The minimum hop counts metric tends to find the route with maximum distance between hops, which may induce the “gray zone” effect, which is discussed in Section 2, and causes severe packet loss.

In the following sections we will discuss different metrics used in selecting routes in wireless ad hoc networks. Furthermore, routing in multi-rate and multi-channel configurations will be surveyed.

3.1 Expected Transmission Count Metric (ETX)

Based on the above experiments and analysis, [24] proposes a new metric to choose routes: Expected Transmission Count Metric (ETX). “The ETX of a link is the predicted number of data transmissions required to send a packet over that link, including retransmissions” [24]:

$$ETX = \frac{1}{d_f \times d_r} \qquad (1)$$

where d_f is forward delivery ratio, by measuring the ratio of packets received by the receiver successfully, d_r is the ACK successful received ratio. These two parameters are measured by link probe packets which are sent over a period τ. The delivery rate from the sender at time t is:

$$r(t) = \frac{count(t - w, t)}{w / \tau} \quad (2)$$

where count(t-w,t) is the number of probes received during window w. The probe packet also has the number of probe packets this node received from each of its neighbors during the last w seconds. Then the receiver can calculate the d_f to the node which sent the probe. In the implementation described in [24], τ=1s, w=10s. The ETX of a route is the sum of the link metrics. From Equation (2) we can see that ETX is based on the delivery ratios, so it can use this information when it selects shorter and higher throughput routes. The routing protocol finds a path that minimizes ETX value.

ETX has been implemented in DSDV and DSR. The simulation results show ETX often finds higher throughput paths than minimum paths, particularly for routes with two or more hops. There are still several issues to be improved on ETX: 1) the size of probe packets is fixed, the prediction of loss ratio will vary with the size of data packets, which results in inaccurate prediction. 2) The frequency of sending probe packets that adapts to different mobility levels, and the overhead it caused. Also in high volume traffic, probe packets may compete with data packets. 3) Working in a network with multi-rate links.

Researchers at Microsoft modify DSR to select a better path by using link quality information [25]. The modified DSR is called Link-Quality Source Routing (LQSR). Three link quality metrics- ETX, per-hop RTT (RTT) [26] and per-hop packet pair (PktPair) [27] - have been implemented in LQSR separately to facilitate comparison between themselves and with minimum hop counts (HOP) metric in a 23-node wireless testbed with TCP transfers. The experimental results show that the ETX metric has the best performance when all nodes are stationary, which is in contrast with the result in [24]. The authors explain the reason is they use TCP traffic and in [24] UDP traffic is used. When a sender is mobile, HOP outperforms all the link-quality metrics, which is contradictory to expectation. The authors contribute it to ETX not reacting to the changes in link quality quickly.

Champaign-Urbana Community Wireless Network combines ETX and Hazy Sighted Link State (HSLS) routing protocol [28] to build a community mesh network. The draft of the ETX protocol specification can be found in [29].

3.2 Routing in Multi-rate Ad Hoc Wireless Networks

The IEEE 802.11 wireless media access standard supports multiple data rates at the physical later. For example, for IEEE 802.11a, the possible data rates are 6, 9, 12, 18, …., 54Mbps and for IEEE 802.11b the set of possible data rates is 1, 2, 5.5 and 11 Mbps. In wireless media, both high speed and long transmission range cannot be achieved simultaneously. In infrastructure based networks, all communication only happens between mobile nodes and the access point. The mobile node can select a transmission rate that works reliably. In ad hoc networks, choosing the data rate becomes more complicated. The routing protocol must make trade-offs between data rate and the distance of the link. Also the path selected by the routing protocol has effect on the congestion level at every node within the interference range of the path. This makes the routing even more complicated.

3.2.1 Automatic Rate Adaptation Protocols

The multi-rate enhancement allows the transmission at several data rates depending on channel quality, but the IEEE 802.11 standard does not specify how to select the rate. There are several auto rate protocols proposed. The Auto Rate Fallback (ARF) protocol [30] is the first commercially available one which was originally designed for Lucent's WaveLAN II device. With ARF, the sender will increase the data rate after consecutive successful transmission and reduce rate after failure. The Receiver Based Auto Rate (RBAR) [31] protocol lets the receiver choose the data rate based on the SNR of the RTS packet. This method can adapt to channel conditions more quickly, but requires modifications to the IEEE 802.11 standard.

The Opportunistic Auto Rate (OAR) protocol [32] opportunistically uses the high quality channel whenever it is available to send multiple back-to-back data packets. OAR depends on two mechanisms: it uses RBAR to access the media and uses the IEEE 802.11 mandated fragmentation field to hold the channel for an extended number of packet transmissions. The key to OAR is that channel coherence times are typically at least multiple packet transmission times. In IEEE 802.11 each node has equal opportunity to send the same number of packets, so that the node transmitting at high speed actually does not gain high throughput if it shares the channel with some nodes at lower transmission rate. With OAR, each node accesses the medium for the same amount of time, so the overall throughput gains are up to 50% compared to RBAR based on NS2 simulations. The implementation of OAR is described in [33].

3.2.2 Medium Time Metric (MTM)

In wireless ad hoc networks with multiple transmission rates, traditional route selection metrics are not suitable.

- Minimum Hop Path: The minimum hop count metric often results in longer links in the route, which corresponds to low channel quality and low transmission rate; also because two nodes of a link are often on the edge of the transmission range, it will suffer from "communication gray zone" problem.

- Shortest Widest Path: It selects the shortest path from the set of paths that have the fastest bottleneck link. This metric works well in wired networks, but in wireless networks, the transmission will interfere a large area.

The simulations in NS2 [34] show the relationship between the throughput across the path and the length of the path: 1) At certain distances low rate links can achieve higher throughput than high rate links, because high rate path may take more hops. 2) Due to spatial reuse, as the path becomes longer, multiple transmissions can take place along the path at the same time. 3) High rate links can achieve high throughput after this distance though more hops needed.

[34] proposes a new route selection metric in multi-rate ad hoc networks: Medium Time Metric (MTM). The authors claim that it can select optimal throughput paths and tends to avoid long unreliable links. MTM assigns a weight to each link in the path, which is proportional to the packet transmission time on that link, and then adds all the weights for the path. Because in IEEE 802.11b, the RTS and CTS packets are transmitted at 1Mbps base rate, for fast links, the portion of MAC overhead increases. Inverse rate weight does not accurately indicate the total consumed medium time. The authors proposes a new link weight as listed in Table 3 [34], compared to the inverse weight. Because the packet transmission time on the link depends on the packet size, here the IP packet size is 1500 bytes. When applying MTM to on-demand routing protocols such as DSR, it will result in the path lasting longer, but mobility may reduce the performance. In [34] the proactive routing protocol DSDV is modified by using MTM as metric instead of hop count. It also uses OAR as lower layer to provide multi rate access and the current communication rate. The primary advantage of MTM is simplicity. It only needs the link rates instead of link utilization which is difficult to detect, partly because OAR already provides some functionality for this. The simulation results show that by combining MTM and OAR, throughput gains of up to 100% to 200% can be achieved over traditional route selection. The shortcomings of MTM are 1) it does not consider packet loss on the link, retransmission will take more transmission time, and 2) a longer path may be caused by the MTM metric, which will increase contention for the medium, finally decrease performance.

Table 3. Rate Based Link Weights

Link Rate (Mbps)	Inverse Weights	1500 byte packet	
		μsec	MTM Weights
11.0	1.00	2542	1.00
5.5	2.00	3673	1.44
2.0	5.50	7634	3.00
1.0	11.0	13858	5.45

Now we look back to [16]. This paper gives one example to illustrate the so called "unintended interactions" in cross-layer design. It compares the performance of Rate Adaptive MAC with original IEEE 802.11, and concludes that the later outperforms the former. Based on the preconditions the authors applied on the comparison, the conclusion is not convincing. The rate adaptation protocol has to combine with the minimum hop metric, which leads to low performance, while IEEE 802.11 only uses 11Mbps data rate to avoid performance degradation.

3.3 Routing in Multi-channel Ad Hoc Wireless Networks

In multi-hop wireless ad hoc network, the performance degrades sharply. Due to the interference from adjacent nodes and neighboring nodes, only one node can transmit at a time. Figure 2 shows an interference example [35]. The IEEE 802.11 standard provides several orthogonal channels (IEEE 802.11b and IEEE 802.11g have 3, IEEE 802.11a has 12); so that multiple communications can happen at the same time to improve the network capacity. There are several researchers recently working on using multi-channels in ad hoc networks. Because the IEEE 802.11 devices can switch channels dynamically, but only work in half-duplex mode, it can only transmit or listen on one channel at a time. Basically we discuss these solutions in two categories: single interfaces and multiple interfaces.

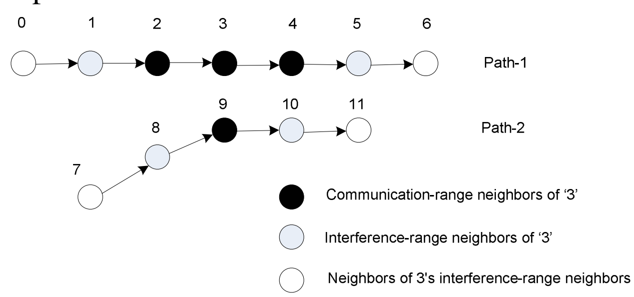

Figure 2. Intra-path and Inter-path interference in a single-channel multi-hop ad hoc network. None of the wireless links shown in the figure can simultaneously operate when node 3 in transmitting to node 4

3.3.1 Multi-channel with Multiple Interfaces

We first look into multiple interfaces. Providing each node with multiple wireless interfaces has some advantages over a single interface:

- Nodes can send and receive simultaneously
- Nodes do not need to synchronize with other nodes for the channel
- Nodes do not need to modify MAC layer protocol, backward compatible
- IEEE 802.11 interfaces are off-the-shelf commodity and the price drops rapidly.

[35] proposes a scheme of centralized channel assignment, bandwidth allocation and routing algorithm. This scheme equips each mobile node with multiple commodity IEEE 802.11 interfaces. The simplest way is assigning each channel to one interface statically as in [36], but it only has limited capacity gain. The main constrains applied to the channel assignment are: [35]

- The number of distinct channels that can be assigned to a node is fixed
- Two nodes that communicate should be on the same channel
- The raw capacity of a channel is fixed
- The total number of radio channels is fixed

One approach called "neighbor partitioning scheme" starts with one node. This node partitions its neighbors into n groups and assigns each interface to each group. The remaining nodes assign channels in the same way but follow the previous settings of other nodes. The problem of this scheme is it uniformly assigns channel across the network and does not consider the load requirement of the flow. Another approach is called "load-aware channel assignment algorithm". Its main idea is to assign channels based on the expected load and the capacity of the channel so that higher traffic load can have more bandwidth. Figure 3 shows the basic flowchart of the scheme [35]. The channel assignment and routing algorithm progress interactively with exploration phase and convergence phase, until the cross-section goodput of the network converges. The cross-section goodput X of the network is: [35]

$$X = \sum_{s,d} C(s,d) \qquad (3)$$

where C(s,d) is the useful network bandwidth assigned between a pair of ingress-egress nodes (s,d). This channel assignment scheme can be applied to any routing protocol. Simulation in NS2 shows that the cross-section goodput can have over 8 times improvement with 2 interfaces per node. The advantage of multiple interfaces is it does not need channel switching and re-synchronization with other nodes.

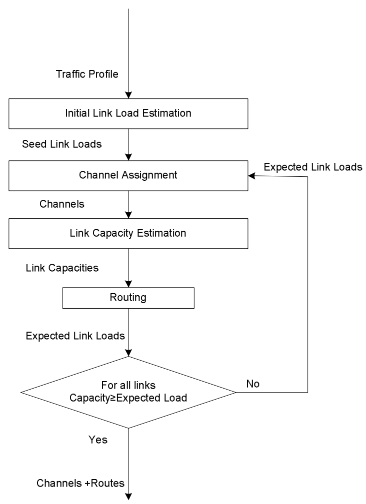

Figure 3. Basic flowchart in multi-channel mesh network architecture

[37] proposes Multi-Radio Link-Quality Source Routing (MR-LQSR) routing protocol for multi-hop wireless network, in which most of the nodes equipped with multiple radios are either stationary or minimally mobile. MR-LQSR is a combination of the LQSR protocol [25] and a routing metric WCETT (Weighted Cumulative Expected Transmission Time). The MR-LQSR's design goals are:

- Consider link bandwidth and PHY-layer loss rate, which both affect packet transmission time.
- The path should be non-decreasing.
- Prefer channel diversity

The routing metric WCETT is:

$$WCETT = (1-\beta) * \sum ETT_i + \beta * \max_{1 \le j \le k} X_j \qquad (4)$$

where β is tunable parameter, 0≤β≤1.
ETT is the expected transmission time:

$$ETT = ETX * \frac{S}{B} \qquad (5)$$

where S is a fixed packet size, B is the bandwidth of the link, which is measured by using the technique of packet pairs [38]. Experimental results show that the packet pair method can estimate the channel bandwidth with reasonable accuracy.

X_j is the sum of transmission times of hops on channel j:

$$X_j = \sum_{Hop\,i\,is\,on\,channel\,j} ETT_i \qquad (6)$$

Equation (4) can be interpreted as a tradeoff between throughput (the 2nd term) and delay (the 1st term). The reason that the ETX metric cannot be used in a multi-channel environment is that it will prefer low rate links and does not prefer channel diversity [37]. The experiments were performed on a 23 nodes wireless testbed. Each node has two different wireless cards. The channels are assigned to the cards statically; one card operated in IEEE 802.11a on channel 36, and the other operated in IEEE 802.11g on channel 10. Both cards used auto-rate. 100 TCP transfers are selected randomly from the node pairs, lasting for 2 minutes. β equals 0.5. The results show that the median throughput using WCETT is 89% higher than ETX and 254% higher than shortest-path routing. Further comparison shows that WCETT can take much better advantage of multiple radios than ETX. On the other hand, WCETT provides less improvement for longer paths than it does for shorter paths due to some problems with TCP which multiple radios cannot address.

Though IEEE 802.11 provides multiple channels at the physical layer, currently the MAC layer is designed under single channel assumption. One of the problems it may cause is so called "multi-channel hidden terminal problem" [39]. This is similar to the hidden terminal problem in a single channel environment, but because of different channels the mobile nodes use, exchanging RTS/CTS message cannot solve this problem. [34] proposes a Dynamic Channel Assignment (DCA) protocol to reserve on-demand a channel. It requires a node to have two wireless interfaces, one channel is used as control channel to transmit control messages like RTS/CTS, and the remaining channels are used for data and ACK transmission. The channel reservation is realized by exchanging RTS/CTS packets. The advantage of DCA is it does not change the IEEE 802.11 MAC protocol, and does not need clock synchronization between mobile nodes. The disadvantage is that channel utilization is not good, because one channel is dedicated to control messages. Also when the number of node increases, the control channel becomes a bottleneck, potentially reducing overall throughput. This is confirmed by simulation results in [39] and [40].

3.3.2 Multi-channel with Single Interface

Using multiple interfaces to achieve multi-channel has the main disadvantages of higher hardware cost and higher energy consumption. The Multi-channel MAC (MMAC) protocol [40] enables mobile nodes to use multiple channels by switching channels dynamically with only one wireless interface. It also can solve the "multi-channel hidden terminal problem" by asking all nodes to listen to a default channel. This protocol is similar to IEEE 802.11 Power Saving Mechanism under the assumption that mobile nodes are synchronized. Time is divided into beacon intervals, and nodes are able to start and finish each beacon interval at the same time. At each beginning of beacon time, there is a small window called ATIM (Ad hoc Traffic Indication Message) window where all nodes listen to a default channel, and then nodes that have packets to send can negotiate the channel assignment with the destination node. Each node maintains a Preferable Channel List (PCL) which records the channel usages in its transmission range, which are categorized into three states, and its preferable channel. The node S that has data to send sends an ATIM packet to the destination D with its PCL. D will select the channel based on its PCL and S's PCL and then sends an ATIM-ACK to D. If S can also select the channel specified in ATIM-ACK, it will send an ATIM-RES packet to D. Then S and D switch to the channel and start RTS/CTS procedure. If S cannot select the channel, it will not send ATIM-RES and wait until the next beacon interval to negotiate with D again, which will waste bandwidth. When multiple nodes start negotiating channels at the beginning of a beacon, a random backoff mechanism is used to avoid collision. Simulations have been done in NS2 with 100 nodes in a 500mx500m area. Each node has one wireless interface with 3 channels, 40 sources and 40 destinations, CBR traffic and 512 bytes packet size. The results from a multi-hop network show that MMAC performs a little bit better than DCA, when network load becomes very high, DCA's throughput drops more quickly than MMAC because of high contention. Larger packets can reduce the number of control messages to alleviate the contention in the control channel in DCA; finally its performance is closer to MMAC. There remain some issues to further improve MMAC, for example: the synchronization between mobile nodes and the overhead it may cause; synchronization between two partitioned networks and improved channel utilization.

Slotted Seed Channel Hopping (SSCH) [41] is another technique proposed recently by Microsoft which allows nodes with one interface to use multi-channel by channel switching. It is a distributed link layer protocol, and does not need tight synchronization between nodes and MAC layer modification. In SSCH, the time allocated to a single channel is defined as a slot, which is 10 ms in the implementation, corresponding to 35 packet transmission times at 54 Mbps. The channel schedule, which describes the node's plan for channel hopping in the future, is compactly represented as 4 (channel, seed) pairs (x_i, a_i), where channel x_i is from 0 to 12, and seed a_i is from 1 to 12. The node will increment each of the channels in its schedule using the seeds after cycling through all the channels in the current schedule:

$$x_i \leftarrow (x_i + a_i) \bmod 13 \quad (7)$$

There are still rare chances that two nodes pick the same seed and never overlap in one channel (with 1 in 20,000 chances). A parity slot is introduced so that every node will switch to this slot after a channel switching cycle. Every node broadcasts its channel schedule in each slot so that nodes can know each other's channel hopping schedule. This is called optimistic synchronization. Schedules are updated in two ways: each node will loosely synchronize the slot's start and finish time with other nodes, or it will overlap another node's schedule if it will send packets to this node. Also, the node will delay channel switching when it is communicating with another node until it finishes. Another strategy called partial synchronization is used for assigning channels, changing schedule and channel congestion prevention.

SSCH is implemented in QualNet [42] for evaluation and comparison performance with IEEE 802.11a. The simulation is done in a 200x200m area. Each node uses single data rate at 54 Mbps. CBR traffic is used with 512 bytes size packet sent every 50 ms. The channel switching delay is set to 80 microsecond. There are simulations run over a number of scenarios. Here are some important results:

- Switching overhead: SSCH takes 500 ms to reach its maximum throughput in a worst-case scenario.
- The synchronization overhead of SSCH is very low, and it can fairly share bandwidth with other flows.
- SSCH normally takes about 500 ms to timeout an absent node.
- In the single-hop case, SSCH outperforms IEEE 802.11a in most cases such as disjoint flows and non-disjoint-flows, except when there are only two nodes in the network.

- In the multihop case, SSCH performs much better than IEEE 802.11a in a multihop chain network when the hop count is bigger than 4, but the throughput already drops from about 12 Mbps to around 2 Mbps. When varying the number of flows, SSCH also achieves significant capacity improvement.
- SSCH was implemented with DSR. Simulations with 100 nodes at slow mobility (maximum is 1m/s) were done. With SSCH, because mobile nodes may communicate in different channels, the sender has to repeat sending broadcast packet in several channels to reach most of its neighboring nodes (the number is 6 in the simulation), which introduces overhead. Because of the switching overhead and broadcast overhead, the average route discovery time is much higher with SSCH (around 0.3 ms for SSCH, much less than 0.1 ms for IEEE 802.11). Also the average route length discovered by SSCH is longer.

There are some topics open to research: the existing routing protocols such DSR do not work well with SSCH, the evaluation of power consumption with SSCH, compatibility with non-SSCH nodes, integrating new routing metrics such ETX with SSCH, etc.

3.4 Multi-Level Hierarchical Routing

The routing protocols we discuss before are called flat routing. In flat routing, the next hop which a mobile node will take to the destination is a physical next hop of this node. In large ad hoc networks (hundreds of mobile nodes), flat routing will cause performance degradation. The main reasons are: firstly, the route's hop count will become bigger in a large scale network, and thus link breakage will happen frequently and end-to-end delay will increase. If some nodes are highly mobile, link failure will become even more severe. Secondly, heavy overhead introduced by routing protocol can consume more network capacity. Thirdly, the routing information about remote nodes can become inaccurate due to the long transmission time. The hierarchical routing protocols are developed to address the network scalability problem.

In hierarchical routing protocols, the network consists of a number of clusters. Each cluster has a cluster leader; traditionally, all the nodes in a cluster are in the direct transmission range of the cluster leader. In some protocols, cluster leaders are more than one hop away, all that is really necessary is for nodes to know how to reach their cluster leader and vice versa. If a node is in the transmission range of more than one cluster leader, it becomes a gateway, and can be used by cluster leaders to relay packets between clusters. Inside a cluster, transmission happens between the cluster head and nodes. Instead of recording a route hop-by-hop, hierarchical routing records a route cluster-by-cluster. Because there may be more than one gateway between two clusters, the route will become more robust.

Numerous hierarchical routing methods have been proposed. A comprehensive review can be found in [43]. Here we briefly present several protocols proposed recently. In hierarchical routing, the overhead and complexity comes from the selection and maintenance of the cluster head. There are several algorithms to select a cluster head, for example, low-ID algorithm [44], weighted algorithm [45] and highest-connectivity algorithm [46]. When one cluster change will cause additional leader changes in the network, this is called rippling effect. The Adaptive Routing using Clusters (ARC) protocol [47] solves this problem by limiting the leadership changes only to occur when one cluster becomes a subset of another cluster. When the number of mobile nodes increases further, multi-hierarchical routing protocols are developed to increase the scalability. In a multi-hierarchical protocol, the heads of the clusters also become a higher level cluster, and the leader of this cluster will be selected. We can represent this scheme as a tree structure as in Figure 4 [47].

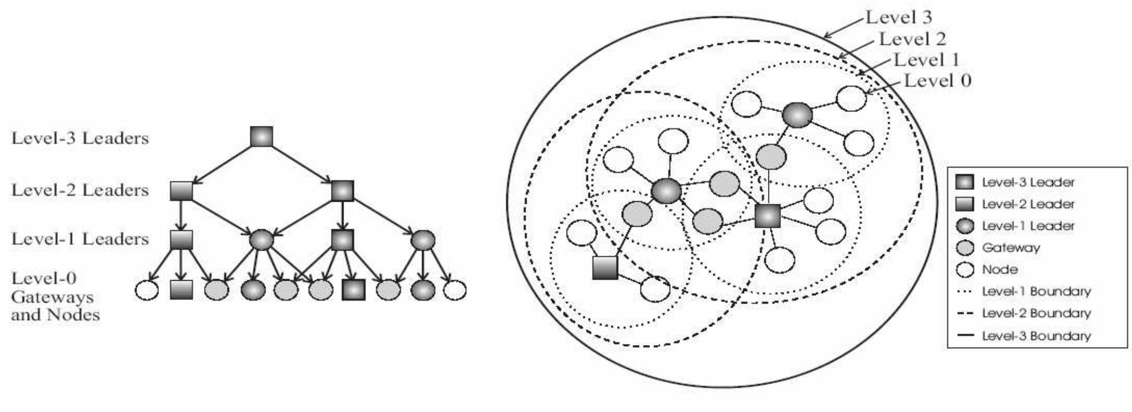

Figure 4. Logical Structure of a Cluster Hierarchy

The Adaptive Routing using Cluster Hierarchies (ARCH) [47] is a multi-level hierarchical clustering protocol which extends ARC. In ARCH, mobile nodes periodically exchange Hello messages between neighboring nodes to build a cluster hierarchy. The adaptive characteristic of ARCH means the hierarchical level can adapt to the changing condition of network. When the network size becomes larger, it will increase the hierarchical level; conversely, it will reduce the hierarchical level. A number of simulations have been done in GloMoSim network simulator with node numbers of between 50 and 750. The simulation shows that few nodes (<2%) are not leader nor gateway, and most of the non-leaders can communicate with more than one cluster leader. This property increases route robustness. Also ARC integrates with AODV by serving as an interface between AODV and IP layer. The cluster leader will process AODV routing packets and provide the next hop (other cluster leader address) and the destination address when it asks a gateway to forward a data packet. The simulations with 100 and 500 nodes show that with ARC, AODV can have higher packet delivery ratio than original AODV (up to 80% in some scenarios).

Safari is another hierarchical routing protocol, which claims to provide large-scale mobile wireless network connectivity and basic network services [48]. It consists of three basic protocols: self organization, scalable routing, and distributed address resolution.

A. **Self organization**: build and maintain a hierarchical structure of the network in a mobile environment. This is done with three basic mechanisms:
    a. Beaconing protocol: in Safari, a subset of mobile nodes, which are automatically self-selected, are called drums. The only special function of drums is to originate beacons. The individual nodes are referred to level 0 cells, and the level 1 cells are also called fundamental cells. Level k cells are grouped into level k+1 cell. Each drum periodically broadcasts a beacon which contains a beacon sequence number, beacon level and coordinate and hop count. The beacon is forwarded by all nodes in its cell and the nodes at the same level that share the same super cell with this node. Higher level beacons are emitted at a lower frequency than lower level beacons. Different from other hierarchical routing protocols, in Safari the

drum at level n transmits a beacon every T_n seconds. This beacon will be forwarded by all nodes within D_n number of hops from that drum, where:

$$\begin{aligned} D_n &= \alpha * D_{n-1} = \alpha^{n-1} * D_1 \\ T_n &= \beta * T_{n-1} = \beta^{n-1} * T_1 \end{aligned} \qquad (8)$$

here α and β are system parameters. D1 is based on the routing protocol in the fundamental cell.

All the nodes store the received beacon in a cache called Drum Ad Hoc Routing Table (DART), which is used for self-organization and routing.

b. Drum level selection algorithm: when a node does not receive a beacon for a timeout period, it will become a drum. There is a series of rules for dynamically changing drums.
c. Membership algorithm: A node runs the membership algorithm after it has run the drum level selection algorithm. Each node at level n will associate with a drum of level n+1 which is the closest one to this node among all the level n+1 drums. Then the node will be assigned a coordinate which is the coordinate of the drum that it associates with. The drum's coordinate is the concatenation of its upper level drum's coordinate and a uniform random number with b bits. The coordinate value will be used in routing.

B. **Scalable routing protocol**: Safari uses a two-step routing process: proactive inter-cell routing and on-demand intra-cell routing. When a node has a packet to send, it will first check the packet's destination coordinate. If their coordinates are match, which means that they are in the same fundamental cell, the intra-cell routing will be used; otherwise, the inter-cell routing will be used
   a. Proactive inter-cell routing: when a node receives a beacon, it saves the direct upstream node's identifier in its DART. Therefore the inter-cell routing can follow the reverse path of the beacon which was sent by the drums where the destination node belongs to. Figure 5 is an example of inter-cell routing. Sender S takes 3 hops of inter-cell routing to reach the fundamental cell that destination D belongs to. [49]
   b. Local route repair in reactive inter-cell repairing: because of node mobility or wireless media fluctuation, the reverse path may be broken. The local route repair protocol uses Route Request and Route Reply mechanism to find an alternate route.
   c. Reactive intra-cell routing: Any ad hoc routing protocol can be used inside the fundamental cell. In Safari, a modified version of DSR is used. Original DSR will flood the Route Request message all over the network. In Safari, the destination will be in the same fundamental cell or with high probability within several hops away from the fundamental cell it used to be. So the number of hop counts of the Route Request is limited in the modified DSR used in intra-cell routing.

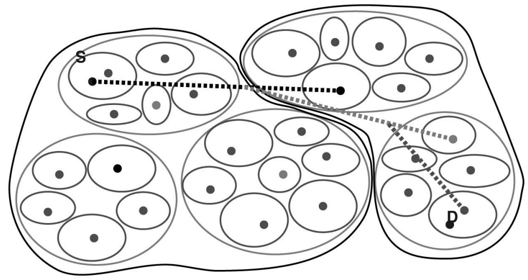

Figure 5 Safari Routing Overview

C. **Address resolution:** Safari uses a distributed way to map a node's identifier to its current coordinate. It is implemented similar to distributed hash table (DHT) used in peer-to-peer networks.

Simulations have been run in NS2 using 2MB base data rate. The number of node varies from 50 to 1000, CBR traffic with 64 byte packets, 4 packets per second. Simulation areas are selected in such a way that the node density corresponds to 50 nodes in a 1000m x 1000m area. The simulation results are:

- Scalability: the packet delivery ratio (PDR) of original DSR drops to about 60% when the number of node reaches 1000, while Safari still has 95% PDR. Also PDR changes little with the increases of traffic load.
- Overhead: when the network size increases, the overhead per node is bounded.
- Mobility: when in 1000 nodes' network, the PDR is over 95% even when all the nodes are mobile. The overhead does not vary with mobility.
- The size of fundamental cell is fixed, independent of network size.

The simulation results of Safari seem promising, which owes to hybrid routing protocols and maintaining an appropriately-sized fundamental cell.

The Hierarchical LANMAR (H-LANMAR) [50] is another multilevel hierarchical routing protocol. The key difference from others is it assumes there are some special nodes in the ad hoc network which are equipped with several powerful, long range radios in addition to the general radios, which are called backbone nodes (BN). The higher level links can be established to connect the BNs to become a backbone network. Recursively we can build a hierarchical network to reduce the long hop counts. H-LANMAR consists of three parts:

- Optimizing the number of BNs: The cluster head in each cluster is elected to transfer traffic between clusters. A small cluster size means small number of mobile nodes and less hop counts, which leads to high throughput. On the other hand, small cell will create more BNs, which in turn reduces the backbone network throughput. Assumes the network is uniformly partitioned and network traffic is uniformly distributed. The total number of mobile nodes is a constant N, and the number of BNs is m, then the optimal number of BN is $\frac{W_2}{W_1}\sqrt{N}+1$, where W_1 and W_2 are the radio bandwidth of the local cluster and backbone network.
- BN deployment: Random Competition based Clustering (RCC) is designed to build cluster. The difference between RCC and other algorithms is that in RCC a node gives up its cluster head position only when it hears another cluster head, thus reduces cluster

instability. RCC also can be extended to hierarchical network of clusters with maximum k hops from the nodes to the cluster head.

- Routing: the routing protocol for the Mobile Backbone Network (MBN) is an extended version of the Landmark Ad Hoc Routing protocol (LANMAR) [51]. The packets are routed to the nearest BN, from where they reach the destination. This greatly reduces the number of hops.

Simulations have been performed in GloMoSim with 1000 mobile nodes in a 3200x3200m area, 30 CBR sources, and node mobility from 0 to 10m/sec. The channel bandwidths are set to 4Mbps for long range radio and 2Mbps for short range radio. The results show that H-LANMAR clearly outperforms AODV and flat LANMAR in terms of delivery fraction, throughput and end-to-end delay. In low mobility, AODV has lower routing overhead, when mobility increases, routing overhead of AODV increases quickly and will be higher than LANMAR and H-LANMAR. On the other hand, LANMAR and H-LANMAR are little affected by mobility.

3.5 Multi-path Routing

Most proposed wireless ad hoc routing protocols are unipath protocols, which only use a single path to send packets to the destination. Because of the dynamic characteristics of ad hoc networks, the wireless links tend to break. A new route has to be found before the source can continue sending packets, which will take time and increase delay. In multi-path routing, multiple routes may be found between the source and the destination nodes. The advantage of the multi-path routing includes:

- Load balancing: distributing traffic over multiple routes, which can alleviate congestion. There are different ways to allocate traffic. For example, a per-connection granularity would allocate all traffic for one connection to a single path; a per-packet granularity would distribute the packets from multiple connections amongst the paths, which may need packet reordering in the destination node.
- Fault-tolerance: when a link breaks, alternative routes still can be used to route the packets. A node disjoint route has a higher degree of fault-tolerant than a link disjoint and non-disjoint route, but it does not ensure transmission independence.
- Higher aggregate bandwidth: multiple paths can be used simultaneously to route data packets.

The disadvantages of multi-path routing protocols, compared to unipath protocols, are their complexity and overhead. They have to discover and maintain multiple paths, and reorder the received packets in some packet scheduling. More information on multi-path routing protocols can be found in [52]. A new routing metric, the route outage probability (ROP), is proposed for channel fading environment with single and multi-path routing [53]. ROP is defined as the probability of packet transmission failure in a route due to channel fading, and can be represented by the average received SNR (Signal to Noise Ratio). In multi-path routing, ROP is used with the multi-route path selection (MRPS) [54] scheme to select a certain number of routes to the destination for the source and intermediate nodes. A node in a route can select the next link which has the best channel state information (CSI) to send the packet among multiple available next hops. The CSI can be obtained by exchanging RTS/CTS packets. The simulation results show that with ROP, MRPS achieves better performance than multi-path and unipath routing. Here are some comments:

- The availability of node-disjoint routes: The advantage of multi-path routing mostly depends on the availability of non-disjoint routes for source and destination pairs. [55] proposes a multi-path extension to AODV called AODV-Multipath (AODVM). Multiple disjoint routes can be found during AODVM's route discovery process. In the simulations in NS2, varying number of nodes (250, 350 and 500) are uniformly distributed in a 2500x2500 m area. The results show that when the hop count from the source to the destination increases, the probability of at least 4 node-disjoint paths decreases. For example, when hop count is 6, the probability is 10% for 250 nodes and around 80% for 500 nodes. Here we may consider the node density. Based on the definition in [48], it is the average number of nodes per transmission range. In NS2, the transmission range is 250 m. So with 250 nodes, the density is $\frac{250}{2500 \times 2500} \times (\pi \times 250^2) \approx 8$ nodes, with 500 nodes, it will be 16 nodes. For conventional simulation parameters, 50 nodes in a 1500x300 m area, the density is about 22 nodes, which is much higher than the 250 node case. We can conclude that in current ad hoc network settings, there is still high probability that multiple node-disjoint routes can be found.

- A recent paper [56] provides a contradictory conclusion to the widely accepted belief that multi-path can significantly improve network balance by numerical analysis and simulations. The reason is when we choose the shortest path, actually the route is very close to the line that connects source and destination in a dense ad hoc network, and the traffic in the center of the network is heavy loaded [57]. Unless a very large number of paths are used (for example 100), using multi-path has a similar effect on load balance as unipath.

3.6 Directional Antennas

Directional antennas can direct radiated power in a certain direction within microseconds. This property can be used in mobile ad hoc network routing protocols to increase spatial reuse. In [58], various combinations of network configuration with directional antenna have been examined. The experimental network has 40 stationary nodes randomly placed in an area which will vary depending on the node density used in the simulation. Here are the results for the simulations:

- CSMA, no Power Control, and Omni-directional Neighbors: simply using CSMA will cause more collisions in neighboring nodes; no power control in directional transmission will introduce more interference, finally no improvement in throughput, but with reduced delay.
- Adding Aggressive Collision Avoidance: this mechanism ensures that the sender will not send a packet to the receiver when the receiver is busy, which reduces many potential collisions. Throughput is increased by 15%.
- Adding Link Power Control: reducing directional antenna transmission power can reduce interference. The throughput can increase up to 28% and delay is dramatically reduced by up to a factor of about 28.
- Using Directional Neighbor Discovery: Mobile nodes periodically send Hello packets to discovery their neighbors. When a node sends Hello messages using directional antenna, the packet will reach further. Thus the throughput can improve up to 118% and a fact of 20 reductions in delay in a low density network.

In [59] a MAC protocol for directional antenna called DiMAC is developed. DiMAC also uses RTS/CTS handshake to reserve a channel and DATA/ACK mechanism. Before sending RTS directionally, the sender will check its table to find the receiver's direction. The receiver is listening in omni mode. When it receives the RTS packet, it will know the sender's direction, and then can send CTS back directionally. The Directional DSR (DDSR) is DSR over DiMAC. When using Qualnet simulator to simulate DDSR, several problems have been identified with using directional antenna:

- If a node has N beam patterns, when it broadcasts a packet, it has to transmit the same packet N times (sweeping), which will produce extra delays. In DDSR, because of sweeping delay the node receiving RREQ message first may reach the destination first, it often selects sub-optimal routes.
- "Deafness" is caused when the sender S is trying to communicate with the receiver R, while R is communicating with another node, then R cannot hear S. Finally S will drop the packet. [60]
- Neighbor discovery and tracing becomes more complex because of nodes' mobility, which causes more overhead

While directional antennas can increase network capacity by increasing spatial reuse, it also can extend the transmission range, which can be used for connecting nodes far apart, and provides more flexibility in routing. On the other hand, the range extension may increase inference due to high effective isotropic radiated power (EIRP) in the transmission direction, which will increase the contention among the mobile nodes, especially in dense networks, reducing the network capacity. [61] proposes the Adaptive Range Control (ARC) mechanism to control the transmission range dynamically. Every node maintains a cache table to record AOA (angle of arrival), reception time and reception power of the last signal from each neighboring node. These parameters are used to clarify the links and determine the local network density. Because interference only occurs when some nodes are trying to transmit, checking the number of active neighboring nodes can help to decide if extending transmission range is feasible. In ARC, the gain is added on the receiver end to extend the transmission, instead of increasing gain in the transmitter, which will not increase interference on neighboring nodes. When a node receives a packet signal power that is lower than the omni antenna's threshold, it will check if the number of nearby nodes is lower than a specified value and the number of nearby nodes in the transmission direction is very low. If yes, it will add this neighbor to its cache table, and extend its transmission range. ARC is implemented in PHY, MAC and network layer (based on AODV), the simulation results show that it can improve the packet delivery ratio up to 9 times in sparse network.

3.7 Other Proposals

There are some other routing algorithms which do not fit into the previous categories but have some interesting new ideas. The Extremely Opportunistic Routing (ExOR) [62] is one innovative unicast routing technique, which takes advantage of characteristics of wireless networks where packets are transmitted successfully with some probabilities. The basic algorithm can be explained by using the simple network diagram as Figure 6 [62]. Suppose node A is a sender to send packets to the destination D. Because at the physical layer, all the transmissions are broadcast over the wireless medium, different nodes receive the packets successfully with different probabilities as indicated in the diagram. These probabilities are assumed to be obtained by some link state information and stored in every node's table. When A sends a packet to D, it will include the next hop node list with priority in the packet. The priority is determined by the minimum hops to the destination. In Figure 6, the list will be "D, C, B". A modified version of 802.11 MAC is used to send ACKs to the sender. Multiple time slots are reserved for the candidate nodes to send ACKs, which contain the highest priority node's ID known to the sender. With modified ACKs, nodes that hear the ACK, which are sent by the node with higher priority than itself, will not retransmit the packet. The result is that the total number of transmission will be reduced significantly (up to 55% less than the best possible pre-determined route). This technique is also very useful for environments with interference like shadowing or fading, where link quality changes frequently and in short time interval. The routes found by a traditional route discovery process only represent the status at that time and will be out-of-date shortly. If we use higher threshold to choose the links, it will result in longer hop counts and longer delay. ExOR provides a new direction to consider routing in a fluctuating environment with distributed and dynamic route selection.

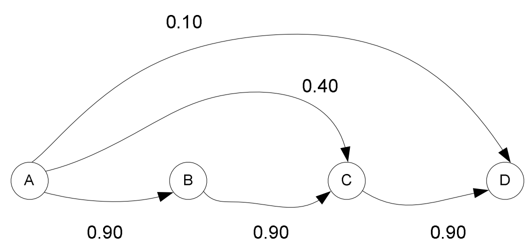

Figure 6. Simple network example, with delivery ratios

MAC-layer anycasting [63] is a forwarding strategy, which allows the MAC to forward packets to any of downstream nodes provided by the network layer based on local channel conditions. It still requires other routing protocols such as DSR to discover and maintain multiple routes. The key is MAC-layer anycasting is more adapt to changing channel condition and can achieve short term optimization. The Fresher Encounter Search (FRESH) algorithm [64] tends to reduce route discovery overhead by searching the intermediate nodes that encounter the destination node more recently than the source node itself. Every node maintains a table which records the "age" that it encountered with its neighboring nodes. The basic FRESH algorithm can be presented as the following pseudo-code:

```
proc FRESH (D) = {
      if (thisnode.ID = D) then {
            replyToSource()
      } else {
            T := prevEncounterAge(D);
            A := findNextAnchor (D, T);
            if (A != D) then
                  notifyNextAnchor(A, D);
      }
}
```

findNextAnchor (D, T) is a network search in concentric ring search manner until it finds an anchor A that has seen D more recently than a time T. In this way, the FRESH algorithm can reduce the route discovery cost by an order of magnitude.

4. Discussion

In the previous sections we reviewed algorithms and new protocols applied to MANET routing protocols in recent years. This section concludes the survey with our evaluation of this work and some comments on promising future avenues of research.

Link quality metrics such as ETX, MTM, WCETT, etc have been proposed as routing metrics to replace the minimum hop count metric, which is widely used by current routing protocols, to select paths in order to increase network capacity. There are still some limitations for applying these metrics to mobile ad hoc routing protocols:

- Though remarkable performance improvement has been achieved by using ETX and WCETT, basically the simulations and experiments are run under mesh networks, where nodes are stationary. When introducing mobility, selecting paths based on these metrics becomes more complicated. These metrics may not indicate whether the two nodes on the link are on the edge of transmission, because if there are few transmissions in this area, the sender can still get good ETX value. If two nodes move apart later, the link will be broken.
- The traffic around a node may change more quickly in a mobile ad hoc network because nodes are joining or leaving neighbor groups, which may increase or decreases the interference. Then the metric may also change. The current algorithms monitor the link quality during data transmission but do not feed this information back to the sender.
- The experiment in [25] using DSR with ETX shows that its median throughput is lower than DSR with minimum hop count metric. The reason is that the ETX metric needs time to become a stable measurement of link quality, on the other hand, the hop count is a simple binary metric and can be used instantly.

In a dynamic environment, we may combine another metric, path longevity, with the metrics above to avoid frequent route switching and to reduce routing overhead. After data transmission starts, some decision will be made when the metric drops to a predefined threshold. There already exists some research on selecting stable routes. [65] [66] provide some metrics to find stable paths. [67] defines a parameter called the stability of the route r, which is:

$$\text{Stability}(r) = (\text{Associativity}(r) / \text{RelayingLoad}(r))$$

The term *Associativity* is the same as defined in the Associativity Based Routing (ABR) [68] protocol; the *RelayingLoad* is the number of routing entries in the routing table of that node. On the other hand, due to the node mobility, some links may finally break during the transmission even considering the route stability at the route discovery process. Link breakage will cause packet delay and more overhead to find a new route. One possible way is to predict the link status and switch to a new route before the link breaks. Some work is required to decide which parameter should be chosen to better predict the link failure and with low overhead before we can apply the prediction algorithm. Ideally, after predicting the link failure, a node has alternative routes to the destination to avoid packet drop and delay. For some routing protocols, such as DSR, nodes store alternative routes in their route caches. These cached routes should be maintained to follow the topology changes. [69] introduces such a mechanism based on a route caching validation probability P_v and local search radius k. The source node will attach a threshold p_t with the RREQ message so that the intermediate nodes can compare their P_v values to decide if the cached routes are “fresh” enough to use.

Some wireless technologies such as high and multiple data rates, multiple channels, and directional antenna have made progress, also many new routing metrics have been proposed in recent years. The researchers intend to use cross-layer design to take advantage of diversity as presented in this survey. The majority of these designs only cross the layer boundaries downwards, i.e. the upper layer protocols directly use parameters provided by a protocol more than one layer below. To route packets in a dynamic ad hoc network environment adaptively, sometimes we may need more proactive ways to change some parameters of mobile nodes, such as transmission power etc, which means the routing protocol may ask lower layers to adjust to the changing environment. A simple example: suppose that two nodes on the link are transmitting data packets, and on the same time they are moving apart. Based on the prediction algorithm, the sender knows in advance that the link will break soon and it unfortunately does not have any alternative route in its cache. While it sends a message back to the source or searches locally to find alternative route, it has to maintain the connection before it can switch to a new link. It can cache the packet, which will increase delay or even cause packet loss when the queue is full; or it can reduce the data rate so that the transmission range will be extended (or increase the transmission power if it already in auto rate mode), it even can use directional antenna to reach the receiver if it knows the its direction.

To design a MANET routing protocol with multiple metrics is a challenge task, especially as the network topology and traffic are changing all the time. We may consider not limiting the mobile nodes to a single predefined routing protocol, instead we let each node decide which protocol to choose based on the environment around it at that time, which has been proposed in [70][71], where it is called active ad hoc routing. Though some hierarchical routing protocols have been proposed to solve the scalability problem, the decision needs to be made what the appropriate or feasible size or node density for an ad hoc network and each of the hierarchical layers should be.

References:

[1] C. E. Perkins, E. M. Royer, and S. R. Das, “Ad Hoc On-Demand Distance Vector (AODV) Routing,” IETF Mobile Ad Hoc Networks Working Group, IETF RFC 3561

[2] D. B. Johnson, D. A. Maltz, and Y-C Hu., “The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks (DSR),” IETF Mobile Ad Hoc Networks Working Group, Internet Draft, work in progress, 24 February 2003

[3] C. E. Perkins and P. Bhagwat, “Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers. In Proceedings of the ACM Special Interest Group on Data Communications (SIGCOMM), August 1994, pages 234-244

[4] T. Clausen, Ed., P. Jacquet, “Optimized Link State Routing Protocol (OLSR)”, Network Working Group, Request for Comments: 3626

[5] E. M. Royer and C.-K. Toh, “A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks”, IEEE Personal Communications Magazine, April 1999, pages 46-55

[6] J. Broch, D. Maltz, D. Johnson, Y.-C. Hu, and J. Jetcheva “A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols”, Proceedings of the IEEE/ACM International Conference on Mobile Computing and Networking (MOBICOM), October 1998, pages 85–97

[7] S. Das, R. Castaneda, and J. Yan, "Simulation-Based Performance Evaluation of Routing Protocols for Mobile Ad Hoc Networks," Mobile Networks and Applications, 2000, Vol. 5, No. 3, pages 179-189

[8] M. Ismail. “Routing Protocols for Ad Hoc Wireless Networks”. M. Sc.(ISS) project, Carleton University, Ontario, Canada, August 2001

[9] S. Shakkottai, T. S. Rappaport and P. C. Karlsson, “Cross-layer Design for Wireless Networks," IEEE Communications Magazine, vol. 41 , No. 10, October 2003, pages 74-80

[10] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, A. Viterbi, “CDMA/HDR: A Bandwidth Efficient High Speed Wireless Data Service for Nomadic Users,” IEEE Communication Magazine, vol. 38 , July 2000, pages 70-77

[11] U. C. Kozat, I. Koutsopoulos and L. Tassiulas, “A Framework for Cross-layer Design of Energy-efficient Communication with QoS Provisioning in Multi-hop Wireless Networks,” Technical Research Report, The Institute for System Research, The University of Maryland, 2004

[12] C. Barrett, A. Marathe, M. V. Marathe, M. Drozda, “Characterizing the interaction between routing and MAC protocols in ad-hoc networks,” International Symposium on Mobile Ad Hoc Networking & Computing, Proceedings of the 3rd ACM international symposium on mobile ad hoc networking & computing, Lausanne, Switzerland, 2002, Pages 92–103

[13] Y. Ko and N. Vaidya. “Location-Aided Routing (LAR) in Mobile Ad Hoc Networks”. 4th Annual International Conference on Mobile Computing and Networking (MOBICOM’98), October, 1998.

[14] H. Gharavi and K. Ban, “Dynamic adjustment packet control for video communications over ad-hoc networks,” The 2004 International Conference on Communications (ICC 2004)

[15] Y.-C. Hu and D. B. Johnson, “Exploiting Congestion Information in Network and Higher Layer Protocols in Multihop Wireless Ad Hoc Networks”, 24th International Conference on Distributed Computing Systems (ICDCS'04), Hachioji, Tokyo, Japan, March 24 - 26, 2004

[16] V. Kawadia and P. R. Kumar, “A Cautionary Perspective on Cross Layer Design.” Submitted to IEEE Wireless Communication Magazine. July 9, 2003

[17] S. Xu and T. Saadawi, “Revealing the problems with IEEE 802.11 medium access control protocol in multi-hop wireless ad hoc networks”, Computer Networks: The International Journal of Computer and Telecommunications Networking, Vol. 38, No. 4, March 2002, pages 531 – 548

[18] H. Ju, I. Rubin, and Y. Kuan, "An Adaptive RTS/CTS Control Mechanism for IEEE 802.11 MAC Protocol", to appear in Proceedings of IEEE Vehicle Technology Conference (VTC), vol.2, April 2003, pages:1469 - 1473

[19] S. Ray, J. B. Carruthers and D. Starobinski, “RTS/CTS-Induced Congestion in Ad Hoc Wireless LANs", WCNC 2003: Wireless Communications and Networking Conference. New Orleans: March 16 - 20, 2003

[20] G. Berger-Sabbatel, F. Rousseau, M. Heusse, A. Duda: “Performance anomaly of IEEE 802.11b,” IEEE The Conference on Computer Communications (INFOCOM) 2003, San Francisco, USA, March 30-April 3, 2003

[21] H. Lundgren, E. Nordström, C. Tschudin “Coping with communication gray zones in IEEE 802.11b based ad hoc networks”, 5th ACM international workshop on Wireless mobile multimedia (WoWMoM 2002), Atlanta, Georgia, September 28, 2002

[22] C. Tschudin, R. Gold, O. Rensfelt, O. Wibling, “LUNAR: a Lightweight Underlay Network Ad-hoc Routing Protocol and Implementation”, Proceedings of Next Generation Teletraffic and Wired/Wireless Advanced Networking (NEW2AN'04)

[23] D. S. J. De Couto, D. Aguayo, B. A. Chambers, and R. Morris, “Performance of Multihop Wireless Networks: Shortest Path is Not Enough”, Proceedings of the First Workshop on Hot Topics in Networking (HotNets-I), Princeton, New Jersey, October 2002

[24] D. S. J. De Couto, D. Aguayo, J. Bicket, and R. Morris, “A High-Throughput Path Metric for Multi-Hop Wireless Routing”, Proceedings of the 9th ACM International Conference on Mobile Computing and Networking (MobiCom '03), San Diego, California, September 2003

[25] R. Draves, J. Padhye, and B. Zill, "Comparison of Routing Metrics for Static Multi-Hop Wireless Networks", ACM Special Interest Group on Data Communications (SIGCOMM), Portland, OR, August 2004

[26] A. Adya, P. Bahl, J. Padhye, A. Wolman, and L. Zhou. "A multi-radio unification protocol for IEEE 802.11 wireless networks". BroadNets, 2004.

[27] S. Keshav. "A Control-theoretic approach to flow control". SIGCOMM, September 1991.

[28] C. Sanivanez and R. Ramanathan, "Hazy Sighted Link State routing protocol (HSLS)" ,BBN Technical Memorandum No. 1301, August 31, 2001

[29] http://wireless.ucimc.org/downloads/ETXprotocol.txt

[30] A. Kamerman and L. Monteban, "WaveLAN-II: A High-Performance Wireless LAN for the Unlicensed Band," Bell Labs Technical J., summer 1997, pages 118-133

[31] G. Holland, N. H. Vaidya and P. Bahl, "A Rate-Adaptive MAC Protocol for Multi-Hop Wireless Networks," in Seventh Annual ACM International Conference on Mobile Computing and Networking (MOBICOM)}, July 2001

[32] B. Sadeghi, V. Kanodia , A. Sabharwal and E. Knightly, "Opportunistic media access for multirate ad hoc networks", International Conference on Mobile Computing and Networking, Atlanta, Georgia, USA, 2002

[33] OAR implementation: http://www-ece.rice.edu/networks/software/OAR/OAR.html

[34] B. Awerbuch, D. Holmer, H. Rubens, "High Throughput Route Selection in Multi-Rate Ad Hoc Wireless Networks", will appear in the First Working Conference on Wireless On-demand Network Systems (WONS 2004).

[35] A Raniwala, K Gopalan and T. Chiueh, "Centralized Algorithms for Multi-channel Wireless Mesh Networks", ACM SIGMOBILE Mobile Computing and Communications Review (MC2R) Vol. 8 , No. 2 April 2004, pages 50-65

[36] P. Bahl, A. Adya, J. Padhye, A, Wolman, "Reconsidering the Wireless LAN Platform with Multiple Radios", SIGCOMM Computer Communication Review (CCR), July 2004.

[37] R. Draves, J. Padhye, and B. Zill, "Routing in Multi-radio, Multi-hop Wireless Mesh Networks", a revised version will appear in ACM International Conference on Mobile Computing and Networking (MobiCom) 2004

[38] S. Keshav. " A Control-theoretic approach to flow control", in ACM Special Interest Group on Data Communications (SIGCOMM) September 3-6, 1991, Zürich, Switzerland

[39] S.-L. Wu, C.-Y. Lin, Y.-C. Tseng and J.-P. Sheu, "A New Multi-Channel MAC Protocol with On-Demand Channel Assignment for Multi-Hop Mobile Ad Hoc Networks," in Int'l Symposium on Parallel Architectures, Algorithms and Networks (I-SPAN), 2000

[40] J. So, N. H. Vaidya, "A Multi-channel MAC Protocol for Ad Hoc Wireless Networks", ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc), Tokyo, Japan, May 2004

[41] P. Bahl, R. Chandra, and J. Dunagan, "SSCH: Slotted Seeded Channel Hopping for Capacity Improvement in IEEE 802.11 Ad-Hoc Wireless Networks", will appear in ACM International Conference on Mobile Computing and Networking (MobiCom) 2004

[42] QualNet: http://www.qualnet.com

[43] C. Perkins, " Ad Hoc Networking," chapter 4, Addison-Wesley, December 2000

[44] C. Chiang, H. Wu, W. Liu and M. Gerla. "Routing in Cluster Multi-hop, Mobile Wireless Networks with Fading Channel". In Proceedings of IEEE Singapore International Conference on Network (SICON), pages 197-211, April 1997

[45] S. Basagni, "Distributed Clustering for Ad Hoc Networks", In Proceedings of the 1999

International Symposium on Parallel Architectures, Algorithms and networks", pages 310-315, Australia, June 1999
[46] M. Gerla and J. T. Tsai, "Multicluster, mobile, multimedia radio network", ACM-Baltzer Journal of Wireless Networks, Vol.1, No.3, pages 255-265, 1995
[47] E. M. Belding-Royer. "Multi-level Hierarchies for Scalable Ad hoc Routing." Wireless Networking (WINET), Vol. 9, No. 5, pages 461-478, Sept. 2003
[48] A. K. Saha and D. B. Johnson, "Self-Organizing Hierarchical Routing for Scalable Ad Hoc Networking", Rice University, Department of Computer Science Technical Report, TR04-433.
[49] A. K. Saha and D. B. Johnson, slides, "Self-Organizing Hierarchical Routing for Scalable Ad Hoc Networking", Mesh Networking Summit 2004, Salish Lodge & Spa, Snoqualmie, Washington, June 23-24, 2004
[50] K. Xu, X. Hong, and M. Gerla, "Landmark Routing in Ad Hoc Networks with Mobile Backbones", Journal of Parallel and Distributed Computing (JPDC), Special Issues on Ad Hoc Networks, Volume 63 , Issue 2, 2002, pages: 110 – 122
[51] G. Pei, M. Gerla, and X. Hong, "LANMAR: Landmark Routing for Large Scale Wireless Ad Hoc Networks with Group Mobility", Proceeding of IEEE/ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHOC) 2000, Boston, MA, August 2000
[52] S. Mueller and D. Ghosal, "Multipath Routing in Mobile Ad Hoc Networks: Issues and Challenges," invited paper in Lecture Notes in Computer Science, Edited by Maria Carla Calzarossa and Erol Gelenbe, 2004
[53] M. Park, J. G. Andrews, and S. Nettles, "Wireless Channel-Aware Ad Hoc Cross-Layer Protocol with Multi-Route Path Selection Diversity", IEEE Vehicular Technology Conference, Orlando, Florida, October 2003.
[54] M. Souryal, B. Vojcic, R. Pickholtz, "Ad hoc, multihop CDMA networks with route diversity in a Raylaigh fading channel," Milcom 2001, Communications for network-centric operations: Creating the information force. IEEE, vol. 2, 2001
[55] Z. Ye, S. Krishnamurthy, and S. Tripathi, "A framework for reliable routing in mobile ad hoc network", Proceedings of IEEE INFOCOM 2003
[56] Y.Ganjali, A. Keshavarzian, "Load Balancing in Ad Hoc Networks: Single-path Routing vs. Multi-path Routing," The 23rd Conference of the IEEE Communications Society (Infocom 2004) Hong Kong, P. R. China, March 7-11, 2004
[57] Y. Ganjali, A. Keshavarzian, "Performance Analysis of Reactive Shortest Single-path and Multi-path Routing Mechanism with Load Balance" IEEE Computer and Communications Societies (INFOCOM) 2003, San Francisco, March 30 –April 3, 2003
[58] R. Ramanathan, "On the Performance of Ad Hoc Networks with Beamforming Antennas", Proceedings of IEEE/ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc), pages 95-105, October 2001
[59] R. R. Choudhury and N. Vaidya, "Impact of Directional Antennas on Ad Hoc Routing", the Eighth International Conference on Personal Wireless Communication (PWC), Venice, September 2003
[60] R. R. Choudhury, X. Yang, R. Ramanathan and N. Vaidya, "Using Directional Antennas for Medium Access Control for Ad Hoc Networks", Proceedings of the ACM International Conference on Mobile Computing and Networking (MOBICOM) 2002, Atlanta, Georgia, September 2002

[61] M. Takai, J. Zhou, and R. Bagrodia, "Adaptive Range Control Using Directional Antennas in Mobile Ad Hoc Networks,". Proceedings of the ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile (MSWiM 2003), September 2003

[62] S. Biswa and R. Morris "Opportunistic Routing in Multi-Hop Wireless Networks", Proceedings of the Second Workshop on Hot Topics in Networks (HotNets-II), Cambridge, MA, USA. November 2003

[63] R. R. Choudhury and N. Vaidya, "MAC-Layer Anycasting in Ad Hoc Networks", ACM SIGCOMM Computer Communication Review, vol. 34 , No. 1, January 2004, pages 75 - 80

[64] H. Dubois-Ferriere, M. Grossglauser, and M. Vetterli, "Age Matters: Efficient Route Discovery in Mobile Ad Hoc Networks Using Encounter Ages", ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc) 2003, Annapolis MD, June 2003

[65] M. Gerharz, C. de Waal, P. Martini and P. James, "Strategies for Finding Stable Paths in Mobile Wireless Ad Hoc Networks", Proceedings of the 28th Annual IEEE Conference on Local Computer Networks (LCN), Königswinter/Bonn, Germany, October 2003

[66] M.l Gerharz, C. de Waal, M. Frank, and P. Martini, "Link Stability in Mobile Wireless Ad Hoc Networks", Proceedings of the 27th Annual IEEE Conference on Local Computer Networks (LCN), Tampa, Florida, November 2002

[67] K. N. Sridhar, J. Lillykutty, and S. Rajeev, "Performance Evaluation and Enhancement of Link Stability Based Routing for MANETs", INTERNATIONAL WORKSHOP ON MOBILE AND WIRELESS NETWORKING (MWN 2004), Montreal, Quebec, Canada, August 15, 2004

[68] C-K. Toh, "Associativity-Based Routing for Ad Hoc Mobile Networks", Wireless Personal Communications: An International Journal, Vol. 4, No. 2, March 1997, pages 103-139

[69] C. Zhao, W. B. Heinzelman, "Adaptive Local Searching and Caching Strategies for On-demand Routing Protocols in Ad Hoc Networks", INTERNATIONAL WORKSHOP ON MOBILE AND WIRELESS NETWORKING (MWN 2004), Montreal, Quebec, Canada, August 15, 2004

[70] C. Tschudin, H. Lundgren, H. Gulbrandsen, "Active routing for ad hoc networks", IEEE Communications Magazine, Vol. 38, No. 4, April 2000, pages:122 – 127

[71] C. Tschudin, H. Lundgren, E. Nordstrom, "Embedding MANETs in the Real World", Personal Wireless Communications (PWC) 2003, Venice, Italy, September 23-25, 2003