Distributed Passive Routing Decisions in Mobile Ad-Hoc Networks

Primož Škraba
Department of Electrical Engineering
Stanford University
Stanford, CA 94305
Email: primoz@stanford.edu

Hamid Aghajan
Department of Electrical Engineering
Stanford University
Stanford, CA 94305
Email: hamid@wsnl.stanford.edu

Ahmad Bahai
Department of Electrical Engineering
Stanford University
Stanford, CA 94305
Email: bahai@stanford.edu

***Abstract*— Communication between nodes in mobile ad-hoc networks is a daunting challenge. The dynamic nature of the environment results in a rapidly changing network topology. This paper presents a cross-layer optimization approach and proposes a routing protocol for achieving minimal delay for energy efficient communication in a dynamic multi-hop ad-hoc environment. The proposed protocol is based on minimizing signaling overhead through state-less step-wise routing decisions made at the receiver instead of the sender. This allows the routing decision to be made with respect to locally available information, which ensures that the information is up to date. The protocol is a receiver contention scheme. The contention is based upon a *time-to-respond* value which is calculated locally at each node through a "delay function". The performance of two delay functions is investigated in detail, which gives an important insight into the effect of physical parameters on the proposed protocol.**

I. INTRODUCTION

Most of the inefficiency in the traditional network stack (OSI model) comes from the overhead each layer generates to estimate the state of the network. In a network with rapidly changing topologies, however, the amount of overhead signaling transmissions and the delays introduced by the various layers prove unacceptable. While networks with rapidly moving nodes obviously have very dynamic topologies, networks with much lower mobility can have unstable topologies due to a changing environment. It has been demonstrated that even in a static network placed in a room, communication can be disrupted by human traffic.

The dedicated short range communication (DSRC) protocol allows for ad-hoc setup of wireless networks consisting of vehicles that participate in the transmission and routing of information such as road condition data to other vehicles that may be interested in such information. Medium access and routing protocols for such ad-hoc vehicular applications need to be designed with particular considerations for the dynamic changes in the network's shape and size. One of the specific issues with such networks is the relatively short amount of time that a participating node may reside in the network. The information that is flowing in such mesh networks towards certain direction or destination needs to be routed in a multi-hop fashion while the network configuration may be changing. The ability to take advantage of the possibly short presence of a vehicle within the network to route the message is then a critical aspect in the design of routing algorithms for these applications. Therefore, it is more critical to be able to place an upper limit on the time between the next hop decision and the data transmission. In addition, a dynamic topology renders scheduling schemes with deterministic access times as impractical due to the overhead required in discovering a node's neighbors and determining the schedule. On the other hand, CSMA-based schemes can only give "average" guarantees, which may result in a situation where by the time a node can access the medium, the network is different enough that the selected next hop may no longer be around.

The Distributed Passive Routing Decision (DPRD) protocol proposed in this paper is a combined MAC/routing protocol, which reduces the overhead from both layers to minimize the per hop delay while maintaining stepwise optimality in a robust fashion. Regardless of the metric used, optimal routing requires current global knowledge. However, in a rapidly changing network this is not possible as by the time all the appropriate information is aggregated to make a routing decision, it is often already out of date. In mobile networks where link quality may change rapidly, the routing decision must be made quickly to take advantage of a good channel. In energy constrained networks, routing must also be done with respect to the nodes' remaining energy. However, aggregating this data changes the state of the network, since a node must expend energy to report its energy state. To avoid the problems associated with data aggregation, routing decisions in DPRD are made by the potential receivers rather than the transmitting nodes. To achieve this, DPRD takes advantage of the broadcast nature of the wireless medium, using the MAC layer to implicitly elect the optimal node as the receiver. While global optimality of the routing path cannot be guaranteed, such optimality is in fact irrelevant due to the changing nature of the network. Instead, routing optimality of the proposed protocol will be superior to that of any other implementation of geographic routing when local parameters are considered.

This paper first presents a brief overview of previous work in section II, followed by a description of the protocol in section III. An analysis of the performance of the protocol is then presented in section IV, followed by specific results for two different delay functions in section V. Finally, conclusions are drawn and possible further work is described.

II. PREVIOUS WORK

There has been a substantial amount of work reported in the area of routing in mobile networks. The authors of [1] and [2] give overviews of protocols for mobile and ad-hoc wireless networks. Both show that in such wireless environments, the unstable nature of the topology of the network requires on-demand protocols. The two most well-known examples of this type of protocol is Ad-hoc On-demand Distance Vector Routing (AODV) [3] and Dynamic Source Routing (DSR) [4]. Because most of the above work is based on the 802.11 standard, there has not been much previous work in the design of the MAC layer for mobile networks.

The proposed protocol in this paper is based upon geographic routing. Greedy geographic routing was first proposed when GPS became available. It is the simplest algorithm and in most situations finds a near optimal path. The greedy algorithm was extended in [5], to include perimeter routing, thus allowing geographic routing to make its way around voids and to escape dead-end local maxima. Geographic routing, however, assumes that nodes are location aware.

Developed independently, two schemes have been proposed recently that embody the same idea as the protocol presented here. Geographic Random Forwarding (GeRaF) [6], [7], and Implicit Geographic Forwarding (IGF) [8], both propose receiver contention schemes. The main difference lies in how the receiver is chosen and in the stress of the analysis. The analysis of GeRaF mostly stresses the energy savings achieved through sleep schedules, while IGF has been developed from a more algorithmic point of view, with a specific implementation proposed in the paper. The contribution of the present paper is that the effect of the protocol on the delay is considered instead of just the energy efficiency performance.

III. PROTOCOL DESCRIPTION

This protocol is based on geographic routing, which uses distributed information to achieve state-less operation. This work attempts to further utilize this distributed information to minimize overhead in the lower layers of the network, most notably in the MAC layer. The dynamic nature of the environment results in an unstable network topology. Thus, routes must be created in an on-demand fashion. As described above, given a destination location, geographic routing finds a route on a hop by hop basis, always routing towards the destination, thereby using locally optimum nodes to route a packet. However, in a highly dynamic network, if parameters such as location and link quality are also taken into account, the exchange of this information between all neighbors incurs a large delay penalty due to the short time during which the information is valid. By using the broadcast nature of wireless communication and moving the routing decision to the receivers, the proposed protocol allows this information to be incorporated without having to gather the data at the sending node.

The decision is made through "receiver contention". The sending node sends out a request to send a data packet (RTS). The packet includes the sender's own location and the location of the final destination of the packet. Based on this information, each receiving node calculates its own optimality and maps this into a delay, τ. After this delay, the receiving node transmits a CTS packet to the sending node, unless another node has responded before. This way, the less optimal nodes do not transmit anything and turn off their radios to prevent overhearing.

The mapping from a measure of optimality to delay will be referred to as the "delay function". Consider some function $f(\theta) \rightarrow \tau$ where θ represents all relevant state information. To ensure that an optimal or near optimal node is selected, the delay function should map the state to delay in a monotonically decreasing fashion with respect to optimality (i.e. the more optimal the state of the node is, the smaller its delay will be). If $f(\theta)$ is bounded to relatively small value, during the time of one transmission, the network topology can be considered roughly the same. Because the topology is always changing, the randomness needed to break deadlocks is already in the system. If the network changes very slowly so that deadlocks could potentially cause severe degradation in performance, a small amount of randomness may be added to the protocol.

Before the performance of different delay functions can be investigated, a quantitative metric of optimality must be found.

A. Optimality in Geographic Routing

The choice of a metric representing optimality depends on the constraints of the system. To simplify the analysis, the smallest set of parameters is chosen. For any one hop using geographic routing, the location of the source node, the location of the candidate receiver node, and the destination of the packet must be known. With a fixed transmission range, the optimal node is given as the neighbor of the source node that is closest to the destination of the packet as shown in Fig. (1.a). The metric of optimality is taken as the distance from a receiving node to the destination, D, normalized over the range $[L-R, L]$, where L is the distance from the source node to the destination and R is the transmission radius.

Given a specific topology, there is always a locally optimum next hop node. If a node knew that it is optimal, it should respond to the source node immediately. Since this would require knowing the full network state, a node must estimate the likelihood it is the optimal node. The likelihood is exactly equal to the probability that there are no nodes in the shaded area, as shown in Fig. (1.b)[1]. The area can be found as

$$A(L,R,r) = \int_{L-D}^{x'} \sqrt{D^2-(x-L)^2}dx + \int_{x'}^{R} \sqrt{R^2-x^2}dx \qquad (1)$$

where $x' = \frac{L^2-D^2+R^2}{2L}$, and $D = \sqrt{L^2+r^2-2Lr\cos(\theta)}$. A closed form expression for $A(L,R,r)$ can easily be found from (1). A two dimensional Poisson process is consistent with a "random" network and is used as the network model to find the probability of optimality. It results in a uniform

---

[1] The assumption of the circular transmission area is an idealization. A probabilistic transmission area would not change the analysis significantly.

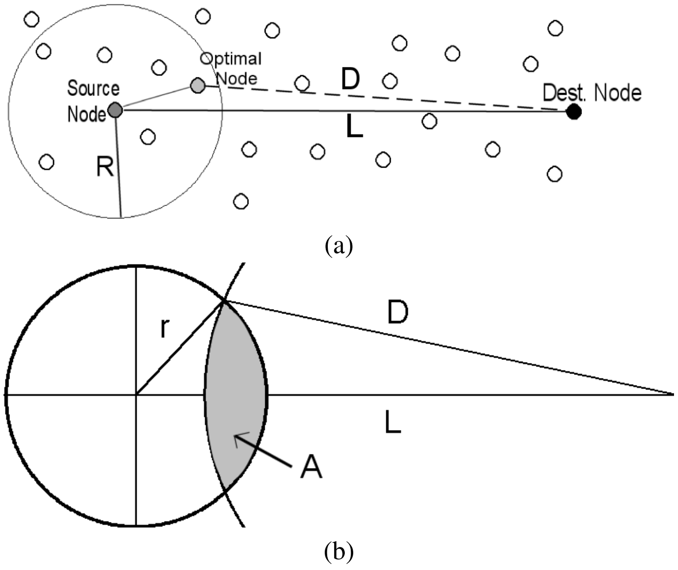

Fig. 1. (a) Example of the one-hop optimal node in a network, the node transmission range is R. (b) Geometric interpretation of receiver position in geographic routing.

distribution for a given number of nodes in a fixed area. For a Poisson process, the probability that there are no nodes in the shaded area in Fig. (1.b) is the probability of first arrival which can be written as

$$\mathbb{P}\{n_j \text{ is optimal}\} = e^{-\rho A(L,R,r)} \tag{2}$$

Note that the probability is given by A and the network density ρ. Implementing the complex calculation for A might not be feasible in a sensor node, but it is possible to approximate it with a much simpler function of $\{L, R, D\}$. For the purposes of present analysis, it is assumed that A or some estimate of A is known.

IV. PERFORMANCE

The main metrics of performance are the average one hop delay and the probability of collision. By modeling both we can examine the tradeoff between these two parameters. Note that regardless of the system, each attempt to find the next hop has a fixed upper bound on the time it takes. Data transmission begins immediately following the next hop decision. Then, the information about the receiving hop being optimal is discarded, and the contention begins again. This is done even if a collision occurs, and ensures that a transmitting node would not try to reach a node that is no longer its neighbor.

Both the one hop delay and the probability of collision depend on the delay function chosen. In sections IV-A and IV-B, performance of two classes of functions is analyzed. To simplify the analysis, the delay function is chosen to be a function of only A. This is because A encapsulates all relevant spatial information about the probability of a node being optimal, and so represents the smallest set of information required for geographic routing. Furthermore, it is assumed that each node is able to calculate or estimate A upon receiving the information from the source node's RTS packet.

In terms of the delay function $f(A)$, the expected value of one hop delay can be found simply by finding $E[\tau]$, which is equivalent to finding the expected value of the delay function:

$$E[f(A)] = \int_0^{A_{MAX}} f(A)e^{-\rho A} dA \tag{3}$$

where $A_{MAX} \approx \frac{\pi R^2}{2}$. Thus the total delay is

$$\begin{aligned} &\tau_{TOTAL\ DELAY} = \\ &\qquad N_{hops}\left(\int_0^{\frac{\pi R^2}{2}} f(A)e^{-\rho A} dA + \tau_{DATA}\right) \end{aligned} \tag{4}$$

where N_{hops} is the number of hops and τ_{DATA} is the time required to send the data.

For the probability of collision the only source of collision that will be considered will be when the CTS packets from two receiving nodes collide. It is assumed that the carrier sense is sensitive enough so that not only can all relevant nodes hear the first CTS, but also it ensures that the neighboring nodes outside the range of the source node do not transmit, because they would sense either the RTS, CTS, or the data transmission.

To find the probability of collision, note that $f(A) : A \rightarrow \tau$, where τ represents the delay from the end of the processing of the RTS. A collision will occur if two nodes calculate two values of τ that are separated by less than $\delta_{CS} + \tau_{PROP}$, where δ_{CS} represents the time required for carrier sense and τ_{PROP} represents propagation delay. Based on this, we can find the following condition to ensure that no collisions occur:

$$\tau_i + \delta_{CS} + \tau_{PROP} < \tau_j \tag{5}$$

where τ_i corresponds to the optimal node and τ_j corresponds to the next optimal node. In (5), τ_{PROP} may be negligible due to the short range of the transmission links. The parameter δ_{CS} is determined by the physical layer. Taking a general function that fulfills the criteria stated above, we wish to obtain an expression for a collision in terms of A rather than in terms of time. For a given A and δ_{CS}, the interval in terms of A is given by

$$g(A, \delta_{CS}) = f^{-1}(f(A) + \delta_{CS}) - A \tag{6}$$

By the independent increment property of the Poisson process, the probability of a collision is given by

$$\begin{aligned} \mathbb{P}\{\text{coll.}|A\} &= 1 - e^{-\rho g(A,\delta_{CS})} \\ &= 1 - \frac{e^{-\rho f^{-1}(f(A)+\delta_{CS})}}{e^{-\rho A}} \end{aligned} \tag{7}$$

Two classes of delay functions are considered in this paper for further analysis: exponential and linear.

A. Exponential Model

To ensure that the exponential delay function is increasing and bounded, the function and its inverse must be of the form

$$f_1(A) = \frac{T_{MAX}\left(1-e^{-sA}\right)}{1-e^{-s\frac{\pi}{2}}} \tag{8}$$

$$f^{-1}(\tau) = \frac{1}{s}\ln\left(\frac{T_{MAX}}{T_{MAX}-\tau(1-e^{-s\frac{\pi R^2}{2}})}\right) \tag{9}$$

where T_{MAX} is the maximum delay allowed by the protocol. Using this we can find the one hop delay

$$\begin{aligned} E[f_1(A)] &= \int_0^{\frac{\pi R^2}{2}} f_1(A)e^{-\rho A}dA \\ &= T_{MAX}\left(\frac{s\frac{1-e^{-\rho\frac{\pi R^2}{2}}}{1-e^{-s\frac{\pi R^2}{2}}}-\rho e^{-\rho\frac{\pi R^2}{2}}}{\rho(\rho+s)}\right) \end{aligned} \tag{10}$$

Trivially, (10) has a minimum at $s = 0$. This is because s must take a positive value (so that the resulting delay is positive). There is a tradeoff between delay and the probability of collision, which is computed next. In this analysis, we use the collision window C, which constitutes the delay over which we wish to study the probability of collision. So in all cases presented here, it can be assumed that $C = \delta_{CS}$. By substituting (9) and (8) into (7) and integrating over all possible values of A, the probability of collision in terms of s and C, is given by

$$\begin{aligned} \mathbb{P}\{\text{coll.}\} =& \frac{1-e^{\frac{\pi\rho R^2}{2}}}{\rho} - \\ & \int_0^{\frac{\pi R^2}{2}}\left(e^{-sA}-\frac{C}{T_{MAX}}(1-e^{-s\frac{\pi R^2}{2}})\right)^{\frac{\rho}{s}}dA \end{aligned} \tag{11}$$

B. Linear Model

For linear functions, there is only one possible form of the delay function

$$f_2(A) = \frac{T_{MAX}}{A_{MAX}}A \approx \frac{2T_{MAX}}{\pi R^2}A \tag{12}$$

$$f_2^{-1}(\tau) = \frac{A_{MAX}}{T_{MAX}}\tau \tag{13}$$

The analysis for this function is straightforward. The delay is given by

$$\begin{aligned} E[f_2(A)] &= \int_0^{\frac{\pi R^2}{2}} f_2(A)e^{-\rho A}dA \\ &= \frac{T_{MAX}}{\pi R^2}\left(\frac{2-e^{-\rho\frac{\pi R^2}{2}}(2+\pi R^2\rho)}{\rho^2}\right) \end{aligned} \tag{14}$$

The probability of collision can be found in the same way as for the exponential function. Again, (12) and (13) are substituted into (7) to find $\mathbb{P}\{\text{coll.}|A\}$. This is integrated over all A to find that the probability of collision is given by

$$\mathbb{P}\{\text{coll.}\} = \left(1-e^{-\rho\frac{\pi R^2}{2}\frac{C}{T_{MAX}}}\right)\left(\frac{1-e^{\frac{\pi\rho R^2}{2}}}{\rho}\right) \tag{15}$$

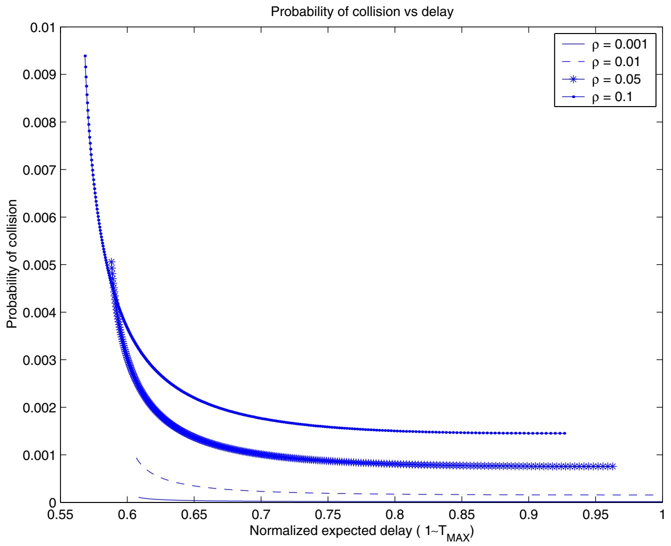

Fig. 2. $\mathbb{P}\{\text{coll.}\}$ vs. delay for several node densities.

V. RESULTS

Based on the above analysis the following observations can be reached. For the exponential function, there is the free parameter s, which allows for different performance characteristics resulting in the operating curves seen in Fig. (2) for a few node densities. It can be observed that by increasing the per hop delay the probability of collision can be reduced marginally. From the plot it can be seen that a Pareto optimal point can be found for minimizing the probability of collision and the average delay. Furthermore, while it can be seen that the performance of the protocol improves with density, the ratio $\frac{C}{T_{MAX}}$ must be chosen appropriately or as Fig. (3) shows, at high densities the probability of collision increases quickly with C.

For the linear function, performance is a function of node density ρ and the ratio[2] $\frac{C}{T_{MAX}}$. As expected, the probability of collision increases with the node density and $\frac{C}{T_{MAX}}$, while the expected delay decreases with the node density and trivially increases with T_{MAX}. Since there is no free parameter, the effect of the different values of $\frac{C}{T_{MAX}}$ at different node densities can be seen in Fig. (4). For increasing values of C/T_{MAX}, the probability of collision can increase rapidly with the node density to unacceptably high levels. Fig. (5) shows the operating curves using the linear delay function at different node densities. An interesting observation is that for a fixed collision window C, by choosing an optimal T_{MAX}, the protocol performs better in low node densities. This result does not consider that a lower density implies a larger probability of a void, which results in a breakdown of the greedy geographic routing and will result in longer delays.

---

[2] In this ratio, it is assumed that C is fixed and T_{MAX} is the design parameter.

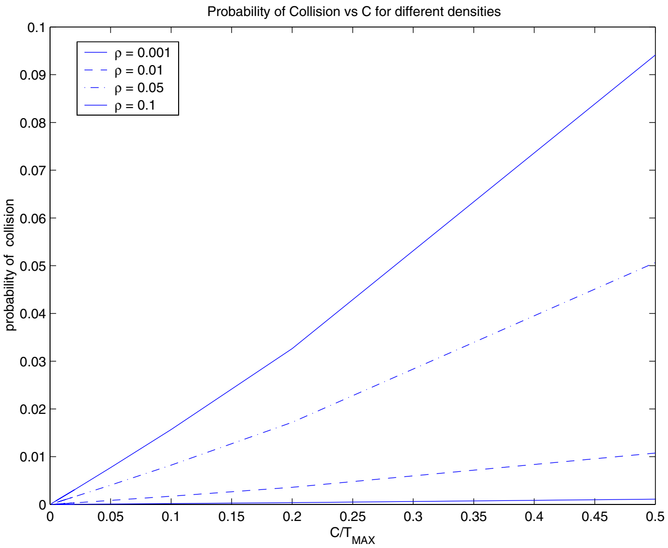

Fig. 3. $\mathbb{P}\{\text{coll.}\}$ vs. $\frac{C}{T_{MAX}}$ for a few node densities.

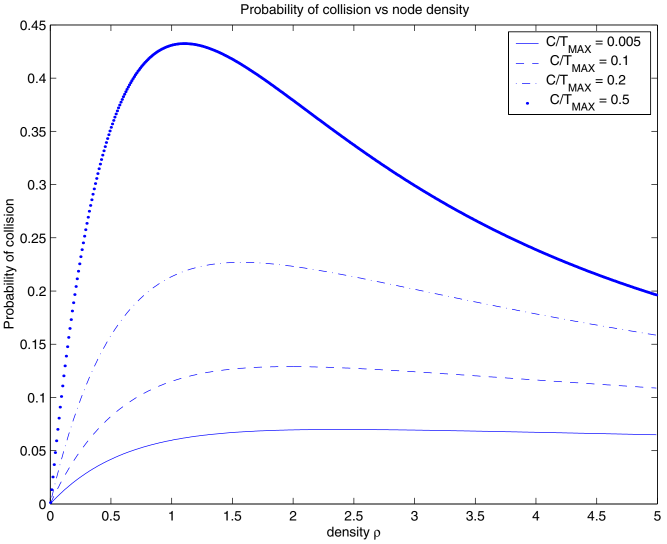

Fig. 4. $\mathbb{P}\{\text{coll.}\}$ vs. node density for $C < T_{MAX}$.

VI. CONCLUSIONS & FURTHER WORK

This paper has presented an analysis of a novel combined routing/MAC scheme that is suited for highly dynamic environments. By integrating the MAC layer with the routing layer, the routing decision and data transmission can be close enough in time that the routing decision is made on current information, even if the environment is rapidly changing. The analysis of two delay functions was presented, but it is expected that in high node densities a more complex function would perform better.

Further investigation needs to be carried out on how to add local parameters to the delay function as well as on analyzing the effect of Rayleigh fading on the probability of collision. Furthermore, obtaining an estimation of network node density would allow for optimal performance. While many such questions need to be addressed, the protocol has many of the desirable properties of deterministic MAC schemes while minimizing the control overhead of making routing decisions.

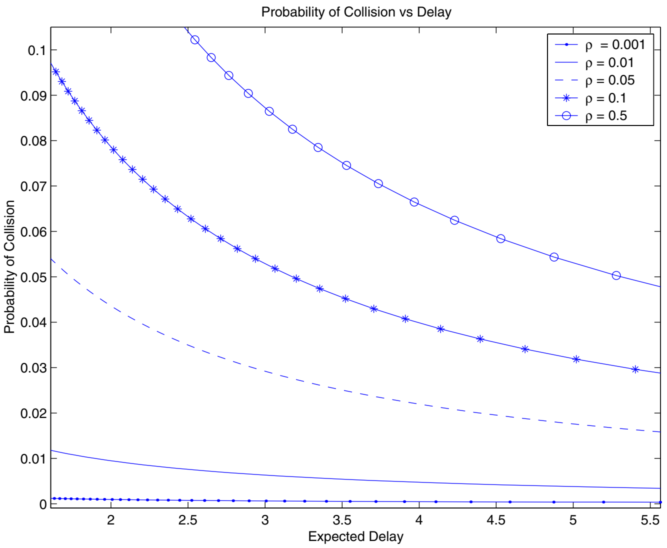

Fig. 5. $\mathbb{P}\{\text{coll.}\}$ vs. delay for a few node densities, varying T_{MAX} assuming $C = 1$.

REFERENCES

[1] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, and J. Jetcheva, "A performance comparison of multi-hop wireless ad hoc network routing protocols," in *Mobile Computing and Networking*, 1998, pp. 85–97. [Online]. Available: citeseer.ist.psu.edu/broch98performance.html

[2] E. Royer and C. Toh, "A review of current routing protocols for ad-hoc mobile wireless networks," *IEEE Personal Communications*, April 1999.

[3] E. M. Royer and C. E. Perkins, "Ad-hoc on-demand distance vector routing," in *Proccedings of IEEE MILCOM*, 1997.

[4] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," in *Mobile Computing*, Imielinski and Korth, Eds. Kluwer Academic Publishers, 1996, vol. 353.

[5] B. Karp and H. T. Kung, "GPSR: greedy perimeter stateless routing for wireless networks," in *Mobile Computing and Networking*, 2000, pp. 243–254. [Online]. Available: citeseer.ist.psu.edu/karp00gpsr.html

[6] M. Zorzi and R. Rao, "Geographic random forwarding (geraf) for ad hoc and sensor networks: multihop performance," *IEEE Transactions on Mobile Computing*, vol. 2, no. 4, pp. 337–348, Oct. - Dec. 2003.

[7] ——, "Geographic random forwarding (geraf) for ad hoc and sensor networks: energy and latency performance," *IEEE Transactions on Mobile Computing*, vol. 2, no. 4, pp. 349–365, Oct. - Dec. 2003.

[8] B. M. Blum, T. He, S. Son, and J. A. Stankovic, "Igf: A state-free robust communication protocol for wireless sensor networks," University of Virginia CS Department, Tech. Rep. CS-2003-11, 2003. [Online]. Available: citeseer.ist.psu.edu/559984.html