Cluster Overlay Broadcast (COB): MANET Routing with Complexity Polynomial in Source-Destination Distance

Luke Ritchie, Hyo-Sik Yang, Andrea Richa, and Martin Reisslein

Abstract

Routing algorithms with time and message complexities that are provably low and independent of the total number of nodes in the network are essential for the design and operation of very large scale wireless mobile ad hoc networks (MANETs). In this paper we develop and analyze Cluster Overlay Broadcast (COB), a low-complexity routing algorithm for MANETs. COB runs on top of a 1-hop cluster cover of the network, which can be created and maintained using, for instance, the Least Cluster Change (LCC) algorithm. We formally prove that the LCC algorithm maintains a cluster cover with a constant density of cluster leaders with minimal update cost. COB discovers routes by flooding (broadcasting) route requests through the network of cluster leaders with a doubling radius technique. Building on the constant density property of the network of cluster leaders we formally prove that if there exists a route from a source to a destination node with a minimum hop count of Δ, then COB discovers a route with at most $O(\Delta)$ hops from the source to the destination node in at most $O(\Delta)$ time and by sending at most $O(\Delta^2)$ messages. We prove this result for arbitrary node distributions and mobility patterns and also show that COB adapts asymptotically optimally to the mobility of the nodes. In our simulation experiments we examine the network layer performance of COB, compare it with Dynamic Source Routing, and investigate the impact of the MAC layer on COB routing.

Index Terms

1-hop clustering, algorithm/protocol design and analysis, message complexity, routing protocol, scalability, time complexity, wireless mobile ad hoc network.

I. INTRODUCTION

Scalable routing is one of the key challenges in designing and operating large scale mobile ad hoc networks (MANETs). In order to ensure effective operation as the total number of nodes in the MANET becomes very large, the complexity of the employed routing algorithms should be low and independent of the total number of nodes in the network. An important consideration in the development of scalable routing algorithms is that the complexity properties of the scalable routing algorithms should be well understood and formally analyzed [1]. While simulations are very useful

---

Please direct correspondence to M. Reisslein.

L. Ritchie, H.-S.Yang, and M. Reisslein are with the Dept. of Electrical Engineering, Arizona State University, Goldwater Center, MC 5706, Tempe AZ 85287–5706, (e-mail: {`Luke.Ritchie, yangkoon, reisslein`}`@asu.edu`, web: `http://www.fulton.asu.edu/~mre`, phone: (480) 965–8593, fax: (480) 965–8325).

A. Richa is with the Dept. of Computer Science and Eng., Arizona State University, (e-mail: `aricha@asu.edu`).

in assessing routing protocols they provide typically only limited insight into the underpinnings and parameter dependencies that govern the algorithm performance. As discussed in more detail in Section I-A, significant progress has been made in recent years in developing and evaluating algorithms and algorithm refinements to achieve scalable MANET routing.

Yet, some key challenges remain in the development and evaluation of scalable MANET routing algorithms. In particular, the existing MANET routing algorithms that have been formally analyzed either:

- incur for the route discovery a total elapsed time or total number of messages exchanged that depend on the overall network size, such as the *total number of nodes in the network or the total diameter (in terms of number of wireless hops) of the network*, see for instance [2–4], or
- make restrictive assumptions about the overall network topology, such as limiting the network density, see for instance [4–6], or assume knowledge of the locations of the nodes at any point in time (location-aided routing), see for instance [7–9].

For these reasons, the analyzed routing algorithms are of limited use for very large MANETs consisting of a very large number of nodes and having very large diameter and no location aid. As detailed in Section I-A, a number of enhancements to the existing routing protocols have recently been proposed to improve their scalability. These enhancements have demonstrated significant potential for improving the scalability in simulations, but have not yet been formally analyzed in the context of the routing protocols.

In this paper we address these two key shortcomings in the state-of-the-art in scalable MANET routing that (i) the existing formally analyzed algorithms do not scale well with the total network size, and (ii) scalability enhancing refinements are formally not well understood. Toward addressing these two points we develop and formally analyze Cluster Overlay Broadcast (COB), a highly scalable reactive routing algorithm for very large MANETs. COB incorporates the recently considered routing on top of clusters mechanism and the doubling radius broadcast mechanism in a judicious manner in a low-complexity reactive routing algorithm.

In brief, our approach is to form a one-hop clustering (cluster cover) of the network and then to perform route discovery by broadcasting route requests over the overlay network formed by the cluster leaders. More specifically, we employ the Least Cluster Change (LCC) algorithm to establish and maintain a clustering structure of the network, whereby a node in a given cluster can reach the leader of the cluster in one hop. When a source node wants to send a message to a destination node, the source node contacts its cluster leader. The cluster leader then floods (broadcasts) route requests over the overlay network of cluster leaders. We employ long-haul transmissions, which have three times the range of the regular (short-haul) transmissions, for the transmissions on the overlay network of cluster leaders. The route requests are broadcast with a doubling radius technique, i.e., the cluster leader first broadcasts the route request with a time-to-live (TTL) of one long-haul transmission hop. If the destination node is reached it responds with an acknowledgement and the route discovery is completed. Otherwise, after a timeout, the source node's cluster leader broadcasts the route request with a TTL of two, then four, and so on.

We formally prove that COB has a route discovery complexity—both in terms of total elapsed time and number of message exchanges—that is polynomially proportional to the *minimum number of hops* between the source node and the destination node and adapts optimally to mobility. More specifically, if Δ denotes the minimum number of short-haul hops from the source node to the destination node, then COB discovers a route with at most $O(\Delta)$ hops (which may include short-haul and long-haul hops). This route discovery takes at most $O(\Delta)$ time and requires the sending of at most $O(\Delta^2)$ messages with COB. We also show that COB requires only a constant amount of storage in each node. These theoretical results, which hold for arbitrary node mobility and node density, build on the constant density of the overlay network of cluster leaders and the doubling radius broadcast. To the best of our knowledge, these results make COB the first MANET routing algorithm for which both the time complexity and the message complexity are polynomial in the minimum hop distance between the source node and the destination node, and independent of the overall network dimensions (total number of nodes, network diameter).

This paper is structured as follows. In the following subsection we review related work. In Section II, we describe the considered model of the MANET. In Section III, we discuss the properties of the algorithm used to maintain the 1-hop cluster cover and prove that the density of the network of cluster leaders is constant for the considered clustering algorithm. We also discuss the system model aspects related to the clustering and the transmissions within and in-between clusters. In Section IV, we introduce and formally analyze the Cluster Overlay Broadcast (COB) routing algorithm. In Section V, we present simulation results for the COB routing algorithm. We summarize our conclusions in Section VI.

A. Related Work

The routing in MANETs has attracted a significant level of interest in recent years, see e.g., [2, 10–16] for overviews. In general, the MANET routing protocols can be classified into *proactive* routing protocols, which maintain routing tables which are consulted when transmitting a packet toward its destination, and *reactive* routing protocols, which find a route on demand, i.e., in response to the generation of a message for a specific destination. Reactive routing protocols are typically more efficient for MANETs with a high level of mobility, see e.g., [16, 17], and are the main focus of our study.

The issue of scalable routing has recently begun to attract significant interest and several studies have formally analyzed and compared the complexity characteristics of the existing routing algorithms, see for instance [2–4, 6, 18–22]. It was found that most of the existing reactive routing algorithms have message complexities that are $O(N)$, where N denotes the total number of nodes in the network. That is, the complexities of the routing algorithms depend on the overall size of the network. It was also found that there are algorithms that have lower complexity, but these generally make restrictive assumptions about the network topology or require location awareness. For instance, hierarchical state routing (HSR) [5, 23] is a proactive routing algorithm that runs on top of a clustering hierarchy and has a message complexity linear in the *average* number of nodes in a cluster and the number of hierarchical levels in the clustering. This makes HSR a low complexity routing algorithm if the nodes are uniformly distributed. In general, if the nodes are non-uniformly distributed the complexity may approach $O(N)$. Also, HSR requires a multi-level clustering hierarchy, which needs to maintain a list of all cluster members in each cluster leader. This structure tends to become costly to maintain for high levels of node mobility. In contrast, COB requires only a simple two-level clustering hierarchy (consisting of regular nodes and cluster leaders) and does not require the cluster leaders to maintain membership lists; COB only requires that each individual regular node knows who its cluster leader is. Examples for routing algorithms that employ location information are location aided routing [8], the greedy perimeter stateless routing [7], and the scalable location update routing protocol (SLURP) [9], which exploit the location information to limit the geographic area over which route requests are broadcast and thus achieve complexities on the order of the geographic broadcast area. COB, on the other hand, does not require any location information and has time and message complexities that are provably polynomial in the minimum source-destination distance and independent of the overall network size.

A plethora of routing algorithms and routing algorithm refinements have been developed and evaluated through simulations. The ad hoc on-demand distance vector protocol (AODV), for instance, which is one of the most prominent reactive routing protocols, has been studied extensively through simulations [17], which have provided invaluable insights into its dynamics and led the development of several refinements. In the adaptive routing using clusters (ARC) approach [24], for instance, the AODV runs on top of a clustering (overlay network) that is maintained with a clustering algorithm that enforces a subset property. That is, a cluster leader is only demoted to regular node status when the cluster has become a subset of another cluster. Further approaches for routing on top of a cluster cover or a set of core nodes have been proposed, see for instance [25–31]. Also, the doubling radius technique has been simulated in the context of AODV [32] and has been found to reduce the complexity. Our work on COB, which also employs clustering and doubling radius broadcast, complements the existing simulation studies of these mechanisms in that we formally analyze these techniques in the context of a routing algorithm. Our theoretical analysis yields fundamental insights into the mechanisms governing the complexities of the routing algorithm. For instance, we find that it is crucial that two cluster leaders are not within the short-haul communications range, as is common with the subset property. The transmission with two different transmission ranges which we employ in COB has been evaluated through simulations in [33].

A variety of other refinements have been proposed which are complementary to our routing algorithm development and analysis. For instance, different mechanisms for flooding the route requests that exploit the mobility of the nodes have been proposed, see for instance [34, 35]. These approaches assume that the nodes either do not move very far or move quite extensively. In contrast we do not assume any specific mobility behavior. The feasibility of routing based on dynamic addresses is examined in [36]. Techniques for further optimizing a route found by a route discovery algorithm are explored in [37].

II. SYSTEM MODEL

We consider a large ad hoc network of mobile wireless nodes (MANET) and let N denote the number of nodes. We consider the problem of unicast routing in the MANET. In particular, we focus on the problems of (i) discovering a route from a source node x to a destination node y, and (ii) delivering a message M from x to y.

We consider a wireless system consisting of homogeneous nodes with the capability to transmit with two different fixed transmission ranges, namely the (normalized) transmission ranges one and three, which can for instance be achieved with power control [33]. Following [33] we use the terminology *short-haul transmission* to refer to a transmission with transmission range one, and *long-haul transmission* to refer to a transmission with transmission range three. For the analytical model we view the short-haul transmission range of each network node as a disk of radius one under Euclidean norm in $\Re^2$ (or $\Re^3$, depending on the particular network application) centered at the node, i.e., each node is represented by a *unit-disk*. Our main motivation for considering the two transmission ranges is to keep our discussions of the routing protocol simple and focused on the main underlying conceptual issues. When only short-haul transmissions are employed, then transmissions between adjacent cluster leaders would need to be forwarded by up to two gateway nodes. We conjecture that the transmissions via the gateway nodes do not affect the asymptotic complexity of the COB algorithm, but would require a more detailed "book keeping" in the clustering. This transmission via gateway nodes is left for future work.

In our analysis we do not assume any specific distribution of the nodes. However, our COB route discovery algorithm—as any other algorithm—can only find a route if the network is connected, i.e., if there exists at least one feasible route from the source to the destination node using only short-haul transmissions with a fixed (normalized) transmission range of one. (We note that an interesting direction for future work is to consider a network scaling scenario where the transmission ranges are adapted to keep the network connected as the number of nodes changes.) We do not assume any specific mobility model in our analysis. We only initially assume, as is reasonable and common, that the mobility of the nodes is on a time scale slower than the route discovery [35].

III. CLUSTERING AS BASIS FOR ROUTING

In this section we present the salient points of the node clustering as it relates to our routing protocol. Node clustering in ad hoc networks has received a significant amount of interest in its own, see for instance [38–45]. Our routing algorithm builds upon some specific properties of the underlying clustering structure. In particular, we employ the Least Cluster Change (LCC) algorithm proposed by Chiang *et al.* [46] for 1-hop clusterings of MANETs. In the clustering we consider only the short-haul transmission range, i.e., each node can reach its cluster leader in one short-haul transmission hop. The choice of the LCC clustering algorithms is motivated by our previous work [47], where we proved that the LCC algorithm is asymptotically optimal or near-optimal with respect to: (i) the number of clusters maintained, and (ii) the cost of an update. More specifically, one would like to minimize the number of clusters maintained, since the smaller the number of clusters maintained, the more efficient the clustering of the network, in the sense that any routing, name lookup, other levels of a clustering hierarchy, or any other network function to be built on top of the 1-hop clustering cover would like to see a network of cluster leaders which is as small as possible (i.e., to have a view of the network which is as simplified as possible). Also, the fewer cluster leaders maintained, the fewer cluster leader changes we expect to see in the network. Thus, we expect a network of cluster leaders which is relatively stable, which is a key property for implementing efficient routing algorithms on top of the 1-hop clustering. We have proven in [47] that the LCC algorithm maintains a 7-approximation on the minimum possible number of clusters, this means that the number of clusters maintained by the LCC algorithm is at most 7 times the minimum possible number in a 1-hop cluster cover of the network, at any point in time.

As far as the cost of an update is concerned, there is a trade-off between the number of clusters maintained and the update cost of an algorithm. For example, we have shown in [47] that if we were able to maintain a minimum 1-hop clustering of the network (note that this problem is NP-hard, and therefore most likely maintaining such a minimum clustering would be infeasible), then the cost of an update may be strictly proportional to the number of nodes N in the network. On the other hand, an algorithm that does not attempt at minimizing the number of clusters maintained could select every node in the network to be a cluster leader, incurring zero update cost. Since this latter approach is equivalent to having no clustering in the network, the best we can hope for is to have constant update cost, keeping each update as "local" as possible. Indeed, we have proven in [47] that the LCC algorithm has an asymptotically minimal update cost, namely (a small) constant.

We proceed by proving a very important property of the LCC algorithm which follows from the fact that no two cluster leaders fall within the short-haul communication range of one another. (Theorem 1 below and the related corollaries also hold for any other clustering algorithm that satisfies the property that no two cluster leaders can communicate directly via a short-haul transmission.) A 1/2-radius disk centered at a node v is a disk of radius 1/2 centered at v. Throughout we employ standard asymptotic notation where a function $g(n) = O(f(n))$ if there exist positive constants c and n_0 such that $g(n) \leq c \cdot f(n)$ for all $n \geq n_0$, and $g(n) = \Omega(f(n))$ if there exist positive constants c and n_0 such that $g(n) \geq c \cdot f(n)$ for all $n \geq n_0$.

Theorem 1: There are at most $O(a)$ cluster leaders whose 1/2-radius disks are fully contained in an area $\mathcal{A}$ of total size a.

Proof: From the property of the LCC algorithm that no two cluster leaders can communicate with one another with a short-haul transmission, it follows that no cluster leader is contained in the unit-disk centered at another cluster leader. Hence, no two $1/2$-radius disks centered at the cluster leaders intersect with each other. Each of these $1/2$-radius disks covers a constant size area, namely an area of size $\pi/4$, of the plane. Thus, if we take an area $\mathcal{A}$ of size a in the network, we can see at most $4a/\pi$ $1/2$-radius disks centered at the cluster leaders which are fully contained in area $\mathcal{A}$. ■

We define the *density* of a network as the maximum ratio of the number of cluster leaders whose

1/2-radius disks are fully contained in an area $\mathcal{A}$ to the total size a of area $\mathcal{A}$, for any area $\mathcal{A}$ on the Euclidean plane. The corollary below follows directly from the theorem.

Corollary 1.1: The network consisting only of the nodes selected as cluster leaders has constant density, namely it has density at most $4/\pi$.

Another corollary, which will be useful when proving the complexity of our COB algorithm, follows from Corollary 1.1:

Corollary 1.2: There are at most $O(r^2)$ cluster leaders in a disk D of radius $r \geq 1$ centered at a cluster leader v.

Proof: The 1/2-radius disks of all cluster leaders contained in D are fully contained in a disk D' of radius $r + 1/2$ centered at node v. The area of D' is $\pi \cdot (r + \frac{1}{2})^2$. From Corollary 1.1, we know that there are at most $4 \cdot (\pi \cdot (r + \frac{1}{2})^2)/\pi = 4 \cdot (r + \frac{1}{2})^2 = O(r^2)$ cluster leaders whose 1/2-radius disks are fully contained in D'. Hence there are at most $O(r^2)$ cluster leaders in D. ■

We remark that if a network is connected when only short-haul transmissions are employed, then the network is also connected when only (i) short-haul transmissions between regular nodes and their cluster leaders as well as (ii) long-haul transmissions between cluster leaders are employed. To see this, note that in a network connected with short-haul transmissions, each node has a neighbor that is no further away than the short-haul transmission range. Thus in a 1-short-haul-hop cluster cover of such a network, where each node is within the short-haul transmission range of its cluster leader, the maximum distance between two adjacent cluster leaders[1] is equivalent to three times the short-haul transmission range, which in turn is equal to the long-haul transmission range.

We also briefly note that in wireless communications the energy consumption increases generally quadratically with the transmission range. A long-haul transmission thus consumes on the order of nine more energy than a short-haul transmission. In our clustering, at least two short-haul transmis-

---

[1] Two cluster leaders are defined to be *adjacent* if there exists a path using only short-haul transmissions via at most two regular nodes between the two cluster leaders.

sions (via at a gateway node) and at most three short-haul transmissions (via two gateway nodes) would be required to communicate from cluster leader to adjacent cluster leader. Hence, the use of long-haul transmissions consumes on the order of between three and 4.5 times more energy than the use of only short-haul transmissions. (We also note that there could be situations in sparse networks where a long-haul transmission can reach a cluster leader that is not adjacent, i.e., more than three short-haul transmissions would be required to reach that cluster leader; in such a situation the use of long-haul transmissions can actually lead to energy savings.) The generally higher energy consumption with the long-haul transmissions can be overcome by forwarding transmissions between adjacent cluster leaders with short-haul transmissions via up to two gateway nodes, which is a direction for future work.

A. System Model: Time Step

In our system model we focus on the network layer and do not consider a particular medium access control (MAC) protocol. We define the *time step* as the maximum time required (i) to conduct a short-haul transmission from a regular node to its cluster leader, or (ii) to conduct a long-haul broadcast from a cluster leader that reaches all regular nodes in the cluster headed by the cluster leader as well as all adjacent cluster leaders, i.e., the cluster leaders within the long-haul transmission range. We assume that the processing of the route requests and acknowledgements in a node takes negligible time (or is accounted for in the time step). We assume that the time step is a constant that is independent from the total number of nodes in the network and the distribution of the nodes in the network. This can be reasonably achieved by employing a mix of time, frequency, and code division multiple access. A cluster leader, for instance, can impose a time division access method for the transmissions from its regular nodes. Also, the transmissions from a cluster leader to its regular nodes and the transmission from a cluster leader to its adjacent cluster leaders (noting that the LCC algorithm ensures that there are no more than 49 such adjacent cluster leaders, as seen by setting $r = 3$ in the proof of Corollary 1.2) can be conducted in different frequency bands or with different CDMA codes [33]. Nevertheless, it is to be understood that in a real wireless network there are in general no absolute (deterministic) guarantees for reaching nodes via wireless transmission within a given time interval, as nodes may experience shadowing or malfunction or other impairments with a non-zero probability. The absolute performance bounds derived for our network model correspond thus in general to probabilistic performance characterizations in real networks.

IV. Cluster Overlay Broadcast Routing

In this section we first describe our COB routing algorithm, to be implemented on top of the clustering structure. We then prove the performance bounds governing this routing algorithm with respect to total elapsed time and total number of messages exchanged. The routing algorithm heavily relies on the constant density of the network of cluster leaders, in order to achieve a polynomial complexity in terms of both time and number of messages exchanged. Also, as we will see, other than requiring an underlying clustering cover of the network at all times, the routing algorithm presented is a purely on-demand algorithm. Thus, in order for this routing algorithm to adapt to mobility in an efficient way, all that is required is that the underlying clustering structure be maintained efficiently upon mobility. We have seen that the LCC algorithm—which is our clustering algorithm of choice—adapts optimally to mobility, namely in $O(1)$ time per event [47].

A. Description of COB Algorithm

In the description of the COB algorithm we let L_z denote the cluster leader of node z, for any node z in the network. We note that each node z in the network needs to be aware of which node is its cluster leader L_z, which each node learns during the establishment of the cluster cover. We also note that COB does not require the cluster leader to maintain a list of the regular nodes in its cluster.

Suppose a node x wants to send a message M to a node y in the network. Node x initiates a route request for node y as described below; all the other nodes in the network follow the COB protocol described below.

1) *Route discovery: Flooding (broadcasting) the route request message on the cluster overlay network:*
    a) Node x starts by sending the message M and destination y to its cluster leader L_x using a short-haul transmission.
    b) Suppose L_x receives the message (y, M) from x at time $t = 1$. Node L_x forwards a route request message (RREQ) of the form $(y, i, 2^i, L_x, x)$, where 2^i is the TTL of the message, at time step $2 \cdot 2^i = 2^{i+1}$, for $i = 0, 1, 2, \ldots$, to all of its adjacent cluster leaders, using a long-haul transmission.
    c) Each cluster leader z receiving a RREQ $(y, i, k, \cdot, \cdot)$ for the *first time* checks whether z is y itself. Otherwise, if $k > 1$, then z forwards a RREQ with TTL equal to $k - 1$ and its own id label, i.e., node z forwards the RREQ $(y, i, k - 1, z, x)$, to its adjacent cluster leaders using a long-haul transmission. Node z keeps the just received RREQ for broadcast round i and discards the stored RREQ from round $i - 1$, if any.
    d) Each cluster leader that still has a stored RREQ $(y, i, \cdot, \cdot, x)$ 2^{i+1} time steps after the receipt of the RREQ, promptly discards the RREQ.
2) *Route discovery: Acknowledging receipt of RREQ and selecting (x, y)-path:*
    a) Node y, upon receiving a RREQ (y, i, k, L_y, x), where L_y is the cluster leader of node y, sends a path acknowledgement message via a short-haul transmission (L_y may be y itself, in which case, we skip the actual sending of the acknowledgement message; also if y is a cluster leader itself, the RREQ message it receives does not contain L_y in its fourth field).

   b) Node L_y, upon receiving a path acknowledgement notice from node y, sends a long-haul transmission path acknowledgement message (ACK) of the form (y, ℓ), where ℓ is the node in the fourth field of the RREQ message stored at L_y. Node L_y also marks itself as ACTIVE(x, y).
   c) Each cluster leader $z,\ z \neq L_x$, upon receiving an ACK message (y, ℓ) checks if $z = \ell$. If so, then z marks itself as ACTIVE(x, y) and sends an ACK message (y, ℓ') via a long-haul transmission, where ℓ' is the node in the fourth field of the RREQ stored at z.

3) *Message transmission:*
   a) If L_x receives an ACK (y, L_x), then L_x
      i) Stops forwarding any RREQ messages relative to M;
      ii) Broadcasts the message (y, M, x) to its adjacent cluster leaders using a long-haul transmission.
   b) Each cluster leader z marked as ACTIVE(x, y), upon receiving a message (y, M, x), forwards the message (y, M, x) via a long-haul transmission. Upon forwarding the last packet carrying the message M, z unmarks itself as ACTIVE(x, y) and discards any ACK or RREQ messages it has with respect to the message M.

Before we analyze the COB routing algorithm we discuss the key steps in more detail. First, note that the successive broadcast rounds $i,\ i = 0, 1, 2, \ldots$, out of cluster leader L_x in Step 1)b) are timed such that the next broadcast round $i + 1$ is only launched if the destination was not reached in the current round i, which is ensured by setting the timeout value for the next broadcast round to twice the TTL field in the current broadcast. A second point to note about Step 1)b) is that for conceptual simplicity, in the described algorithm, a given cluster leader uses only long-haul transmissions and these long-haul transmissions are used to reach both the regular nodes around the cluster leader as well as the adjacent cluster leaders. In particular, broadcast round $i = 0$ reaches all the regular nodes around cluster leader L_x (i.e., the regular nodes for which L_x is the cluster leader) as well as the cluster leaders adjacent to (within the long-haul transmission range of) L_x. (Some regular nodes from the clusters adjacent to L_x may also be reached, but this is not relevant, as explained shortly. Also, reaching the adjacent cluster leaders is only a side effect of using the long-haul transmission. All we need to achieve in round $i = 0$ is to reach the regular nodes around L_x.) Broadcast round $i = 1$ reaches again all regular nodes around L_x as well as all the regular nodes around the cluster leaders that are adjacent to L_x (and in turn their adjacent cluster leaders).

Instead of the using only long-haul transmissions, the broadcast rounds may be conducted with a mix of (i) short-haul transmissions to reach the regular nodes around a given cluster leader, and (ii) long-haul transmissions to reach the adjacent cluster leaders. These two types of transmissions can be conducted in different frequency bands or with different CDMA codes, similar to [33], to simplify the medium access control. When employing this mix of short- and long-haul transmissions the broadcast rounds proceed as follows. In round $i = 0$, cluster leader L_x broadcasts the RREQ with a short-haul transmission to all of its regular nodes. In round $i = 1$, cluster leader L_x broadcasts the RREQ again with a short-haul transmission to its regular nodes. In addition, L_x broadcasts the RREQ with a long-haul broadcast to its adjacent cluster leaders, which in turn broadcast the RREQ with short-haul transmissions to their regular nodes, and so on in the following rounds.

As an additional refinement we can skip the repeated short-haul transmissions, i.e., in round $i = 1$ we skip the short-haul transmission to the regular nodes around L_x which have been reached in round $i = 0$. In general, with this additional refinement only the cluster leaders reached for the first time in a broadcast round, i.e., the cluster leaders receiving route requests with TTL value $k = 2^{i-1},\ 2^{i-1} - 1, \ldots, 1$, forward the RREQ with a short-haul transmission to their regular nodes. We note that using the mix of short- and long-haul transmissions and the described additional refinement do not affect the asymptotic complexity of COB (aside from affecting the involved constants), as analyzed in the following section. However, they tend to simplify the medium access control [33].

We remark to Step 2)a) that a regular node y processes an RREQ as detailed in Step 2)a) only if the RREQ is received from the node's cluster leader L_y. RREQs directed to y that are received from other cluster leaders (e.g., from long-haul transmissions) are ignored.

B. Analysis of COB algorithm

The *message complexity* of a given distributed network algorithm is given by the number of unit-size packet transmissions throughout the execution of the algorithm. The *time complexity* of a distributed algorithm is given by the total elapsed time during the execution of the algorithm. We let the $|M|$ denote the size of a message M in number of packets.

Theorem 2: (a) Route Discovery: The time and message complexity of the route discovery part of the COB routing protocol are $O(\Delta)$ and $O(\Delta^2)$ respectively, where Δ is the minimum hop distance between the source node x and destination node y.

(b) Message transmission: The time and message complexity of the actual message transmission from x to y are both asymptotically optimal, i.e., $O(\Delta|M|)$.

(c) Storage complexity: The COB routing protocol only requires a constant amount of storage space at each node in the network, which will be released once the (x, y)-routing is complete.

Proof: (a) Route discovery: We first prove the complexity of the route discovery part of the COB algorithm. Suppose node y was first reached during the ith broadcast round originated at node L_x. Hence the distance from L_x to L_y must be at most 2^i hops (only cluster leaders within 2^i hops from L_x are reached during the ith broadcast). The broadcast rounds out of node L_x will end as soon as an ACK is received by that node. The RREQ message that first reached node y must have been sent before the ith round was completed, i.e., at a time $t \leq 2^{i+1} + 2^i = O(2^i)$, since the ith broadcast commences at time step 2^{i+1} and takes at most 2^i time steps to complete. The ACK sent out of node L_y must have been sent at time $t+1$. Any ACK sent by the algorithm goes from a node reachable from L_x in h hops to a node reachable from L_x in $h-1$ hops (note that each ACK has a specific node it is trying to reach, namely the one given by the "parent" field, i.e., the fourth field, in the RREQ stored at the node sending the ACK). Thus, since L_y is reachable from L_x after $t' \leq 2^i$ time steps of the ith round, it will take at most 2^i time steps for the ACK originated at L_y to reach L_x. Putting all these costs together, the route discovery takes at most $2^{i+1} + 2^i + 2^i + 2 = O(2^i)$ time steps (the constant additive term comes from the fact that there may be two additional communication steps between L_y and y).

Since we know that y was not reached in the $(i-1)$th round, the long-haul distance between L_x and L_y must be at least 2^{i-1} long-haul hops, implying that the short-haul distance between L_x and L_y must also be at least 2^{i-1} short-haul hops. Since any (L_x, L_y)-short-haul path of the form $L_x, x, \ldots, y, L_y$ is a candidate path for being the path between L_x and L_y with the smallest possible number of short-haul hops (which, as we have seen, must be longer than or equal to 2^{i-1}), the short-haul distance Δ between x and y has to be at least $2^{i-1} - 2 = \Omega(2^i)$. Hence, the route discovery of the COB algorithm takes time which is linearly proportional on Δ, i.e., $O(\Delta)$.

We now prove that the message complexity of the route discovery is $O(\Delta^2)$. In the ith broadcast round, each cluster leader reached in this round sends at most one RREQ. All cluster leaders reached in the ith broadcast round fit into a disk of radius $O(2^i)$ and, thus, by Corollary 1.2, there are $O(2^{2i})$ such cluster leaders. Hence, $O(2^{2i})$ messages are sent in round i. Hence the total number of messages sent in all rounds of broadcast is $O(\sum_{i=0}^{\log \Delta} 2^{2i}) = O(\Delta^2)$. Since $O(\Delta)$ ACKs are sent in the route discovery phase, the message complexity follows.

(b) Message transmission: The message transmission phase only involves the nodes in the selected path from L_y to L_x and each node in this path takes one time step to forward each packet of M to the next node in the path. We have seen that the selected path from L_x to L_y (and hence the extension of this path that goes from x to y) has $O(\Delta)$ hops. Hence the message and time complexity of the message transmission phase are both $O(\Delta|M|)$.

(c) Storage complexity: It remains to show that the COB routing protocol only requires a constant amount of storage space at each node in the network. Note that at any time during the execution of the algorithm, each node (more specifically, each cluster leader) stores at most one RREQ and has at most one ACTIVE(x, y) mark, which are all of constant size. Also, the COB algorithm does not use and therefore does not need to maintain any cluster membership information at the cluster leaders (neither does the LCC algorithm): The only information necessary for the COB algorithm to work is that each node z knows who its cluster leader L_z is, which takes only a constant amount of space. All the ACTIVE(x, y) marks are erased as the message M is indeed transmitted from L_x to L_y, and any RREQ is released after all broadcasts from L_x are guaranteed to have terminated.

■

Corollary 2.1: For fixed size messages, the overall time and message complexity of the COB algorithm are $O(\Delta)$ and $O(\Delta^2)$ respectively.

Proof: Adding up the asymptotic complexities of the route discovery and message transmitting phases in view that $|M| = O(1)$ gives the result in this corollary. ■

Theorem 3: The COB routing protocol adapts asymptotically optimally to the mobility of the nodes.

Proof: The COB algorithm is a purely on-demand algorithm provided we always maintain a clustering cover of the network according to the LCC algorithm. Thus any updates upon mobility of the nodes must only be taken care of at the clustering level. In [47], we have shown that the update cost (total elapsed time of an update) of the LCC algorithm is $O(1)$, which is asymptotically optimal. The number of messages exchanged during an update of the clustering structure is linearly proportional to the number of nodes left uncovered due to the mobility of a node—i.e., if a node v moves and if as a result k nodes are not covered by any cluster leader after the move (and possible demotion of v as a cluster leader), then at most $O(k)$ messages will be sent in order to fix the LCC clustering cover. Note that no deterministic algorithm can have a better message complexity for creating a clustering cover of k nodes. ■

Combining the results in Theorems 2 and 3, we have that the COB algorithm is the first routing algorithm for a MANET of homogeneous nodes that under the unit-disk model adapts optimally to mobility and that has time and message complexities both polynomial on the distance between source node and destination node.

We also note that the route found by the COB algorithm is free from loops, which follows from the fact that an ACK goes from a node reachable from L_x in h hops to a node reachable from L_x in $h-1$ hops, as noted in the proof of Theorem 2(a). We furthermore note that the delay incurred with COB is at most five times larger than the delay with naive flooding over the network of cluster leaders. To see this consider a cluster leader L_y of the destination node that is 2^i+1 long-haul hops from the cluster leader L_x of the source node. With our doubling radius broadcast, L_y is reached in broadcast round $i+1$, which commences at time 2^{i+2}, and in this broadcast round it takes 2^i+1 time steps to reach L_y. Thus, the total delay to reach L_y is $2^{i+2}+2^i+1$ time steps, which is $(5+1/2^i)/(1+1/2^i)$ times larger than the delay with naive flooding. We believe that this larger delay is a reasonable trade-off for achieving a bounded message complexity of $O(\Delta^2)$.

C. *Energy-fair COB*

In COB the cluster leaders conduct all the inter-cluster communication and do so using long-haul transmissions, which consume more energy than short-haul transmissions, as outlined at the end of Section III. This may lead to unfairly high energy consumption in nodes that act as a cluster leader for long periods of time. To address this problem we propose a slight modification of the COB algorithm in order to obtain an energy-fair COB algorithm, i.e., an algorithm which aims at a fair usage of energy at all nodes.

The energy-fair COB algorithm is implemented as follows: Instead of using the ID of a node as the tie-breaker in the clustering algorithm, we use the remaining energy level of a node as the tie breaker. Thus, when nodes compete in a local region for becoming cluster leaders, the one with the highest remaining energy level will win (if there is still a tie between nodes with the same remaining energy level, we can break this tie using the unique node IDs). Whenever the power level of a cluster leader node L_z drops below half of the power level of L_z at the time it was elected cluster leader, then L_z demotes itself as a leader node and starts a local update on the clustering of the network. As discussed earlier, this local update will only take $O(1)$ time. Note that while the power levels of the nodes in the network remain reasonably large, a cluster leader will remain as cluster leader for a significant amount of time and the local clustering updates due to energy drops at the leader nodes will not be frequent. It is only when the network comes to a very low energy level that there will be a significant overhead due to frequent cluster leader swaps. However, at this point, the network is basically at "the end of its life" since the remaining energy at all nodes is indeed coming to an end. We note that similar approaches have been proposed in the context of clustering protocols, for instance in [48, 49].

The energy-fair COB algorithm is fair in terms of energy usage to the nodes in the network, in the sense that the nodes in a local neighborhood tend to converge to a scenario where the energy levels of all the nodes fall in between α and $\alpha/2$, for some energy level α. If the network communication patterns are uniform along the different regions of the network, then we also expect this energy level α to be roughly the same for the entire network.

We note that there are situations in which the energy starvation of a node is unavoidable: Suppose there is a region R of the network with very low density (e.g., if R is a unit-disk and there are only a small constant number of nodes in R), and suppose this region is a "bottleneck region" in that it provides the only bridge between large, densely populated parts of the network. Then no matter how we elect cluster leaders in R, we expect the energy consumption at the nodes in R to be much higher than that of the rest of the network. This is because there are only few nodes in the bottleneck region R that can alternate in performing the role of cluster leader.

V. SIMULATION RESULTS

In this section we present simulation results to illustrate the performance of the COB routing algorithm. We examine three different aspects of the COB algorithms in the simulations, namely (i) the network layer performance of COB, (ii) the comparison of the network layer performance of the well-known Dynamic Source Routing (DSR) algorithm with COB, and (iii) the impact of the MAC layer on the COB performance.

A. Network Layer Performance of COB

Our network layer simulation setup is similar to the route discovery evaluation setup employed in [35] in that we evaluate COB only with respect to the mobility process and the size of the network. In particular, we consider an idealized model of the MAC layer where transmissions reach their destinations within one time step, as defined in Section III-A, and we simulate the route discovery sequentially. This ensures that we measure the network layer performance of COB, in isolation from any positive or negative effects of the MAC layer or cross-traffic.

We conduct simulations for two scaling scenarios: (i) a node density scaling scenario, where the area of the network is a square of fixed size R = 500 m by R = 500 m, and the number of nodes N in the network is varied, and (ii) a network diameter scaling scenario, where we jointly scale up the number of nodes N in the network and the diameter of the network area. In particular, in the diameter scaling scenario we consider the configurations: N = 250 nodes in 125 m by 125 m square area, N = 500 nodes in 250 m by 250 m square area,..., N = 4000 nodes in 2000 m

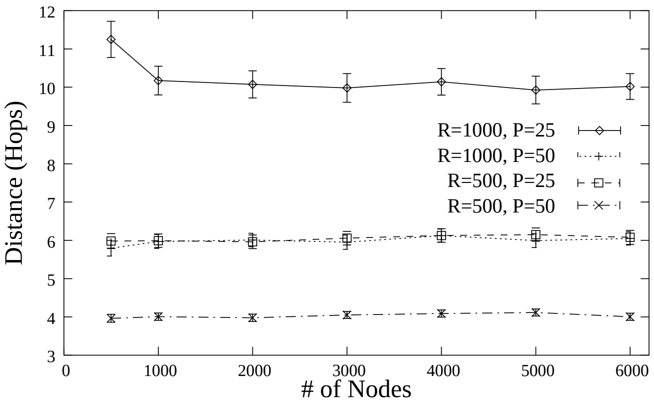

Fig. 1. Node density scaling: Length of discovered route as a function of number of nodes N for different (fixed) network areas of $R \times R$ m^2 and short-haul transmission ranges P m

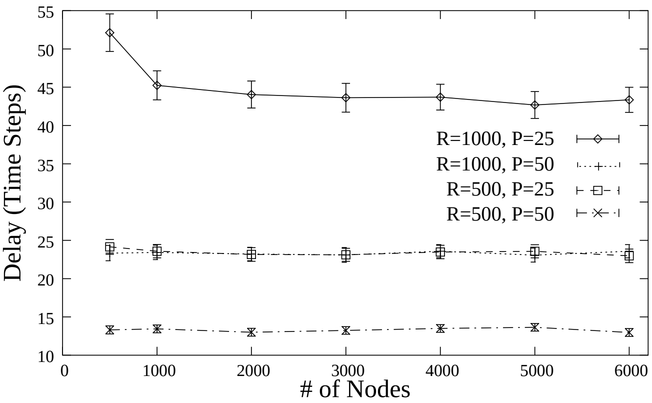

Fig. 2. Node density scaling: Delay for a route discovery as a function of number of nodes N for different (fixed) network areas of $R \times R$ m^2 and short-haul transmission ranges P m

by 2000 m square area. The goal of the diameter scaling is to investigate the performance of the COB routing algorithm as the shortest (short-haul) hop distance Δ between source and destination increases. Clearly, it is computationally prohibitive to find the true shortest route, but it is reasonable to assume that Δ scales approximately linearly with the diameter of the network. Throughout, we consider the two short-haul transmission ranges P = 25 m and P = 50 m, and corresponding long-haul transmission ranges of 75 m and 150 m. We conduct simulations for both the random walk (RW) and the random waypoint (RWP) mobility models with a mobile speed of 10 m/sec, the pause time for the random waypoint mobility model is 10 seconds. Since a route discovery takes typically on the order of tens or hundreds of milliseconds with the idealized MAC layer, whereas changes in the cluster cover due to node mobility take place on the time scale of typically tens of seconds, we approximate the node positions as static during a given route discovery. The practical deployment of the COB routing protocol would of course require that node changes in the cluster cover that affect an ongoing route discovery are properly recovered from.

In the simulations we consider the COB algorithm employing only long-haul transmissions in the broadcast rounds, as detailed in Section IV-A. We conduct sequentially several stochastically independent route discoveries between randomly uniformly chosen source and destination node pairs and collect statistics on the number of messages transmitted and the time elapsed for the route

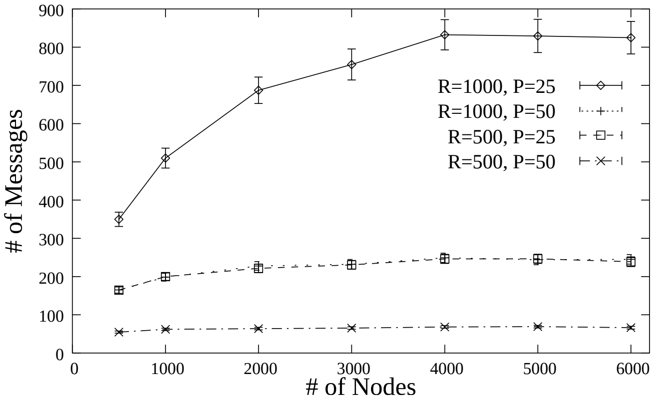

Fig. 3. Node density scaling: Number of message transmissions for a route discovery as a function of number of nodes N for different (fixed) network areas of $R \times R$ m^2 and short-haul transmission ranges P m

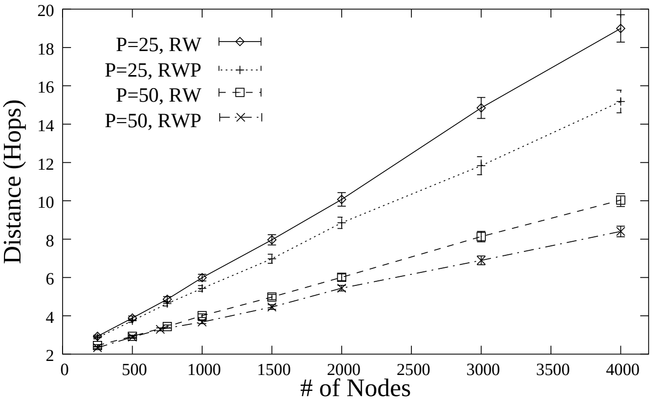

Fig. 4. Diameter scaling: Length of discovered route as a function of number of nodes N with proportional ($R \sim N$) network area of $R \times R$ m^2 for different (fixed) short-haul transmission ranges P m and mobility patterns (random walk (RW) and random waypoint (RWP))

discovery. We also collect the statistics on the number of hops of the route found by the COB algorithm. We continue each simulation until the 95% confidence intervals are smaller than 10% of the corresponding sample means of the measures of interest.

In Figs. 1, 2, and 3 we consider the node density scaling scenario and plot the delay for a route discovery (in time steps) and the number of message transmissions per route discovery. The plotted results are for the random walk mobility model. We observe from Fig. 1 that for a 500 m by 500 m square and a short-haul transmission range of 50 m, the discovered route has on average four hops. Typically such a four hop route consists of one short-haul hop to go from regular source node x to its cluster leader L_x, then two long-haul hops to reach the cluster leader L_y of the destination node y, and one more hop to go from L_y to y. Generally, we observe that the length of the discovered route and the delay for the route discovery do not change significantly as the number of nodes N in the fixed network area increases, i.e., as the node density increases. An exception is the 1000 m by 1000 m network with a short-haul transmission range of 25 m, where the route length and delay decrease as the number of nodes increases from 500 to 1000. This effect is due to “uncovered” areas in the network of 500 nodes, that are not bridged by the short radio transmission range of the relatively few nodes and require a route around the uncovered area, i.e., the routes tend to be more crooked and less straight in this scenario. In any case, Theorem 2(b) ensures that the discovered route is asymptotically linear in the shortest possible route.

From Fig. 3 we observe that the number of messages transmitted for a route discovery generally tends to initially increase and then level off as the node density increases. This effect is most pronounced for the network with the large area and the small transmission range. This effect is due to the initially increasing number of clusters as the network becomes more populated. Once the entire network area—or more precisely the entire disk centered at a given source node with radius required to reach the given destination node with the doubling radius technique—is covered by clusters there is no further increase in the number of messages. Theorem 2(a) guarantees that the number of messages is at most quadratic in the shortest source to destination hop distance irrespective of the overall network size.

In Figs. 4, 5, and 6 we consider the diameter scaling scenario for the random walk (RW) and random waypoint (RWP) mobility models. We observe from Figs. 4 and 5 that the length of the discovered routes and the route discovery delay increase linearly as we jointly scale up the diameter and number of nodes in the network. Also, we observe from Fig. 6 that the number of messages transmitted for a route discovery appears to increase quadratically with the diameter of the network, which in turn gives a good indication of the shortest source-destination hop distance Δ. We note that in our simulation set-up, the source node and destination node are drawn uniformly randomly on the network area. Thus with expanding network area, the source and destination node are increasingly further apart, giving rise to the observed scaling behaviors of the delay and message complexity, which reflect our theoretical results (see Theorem 2(a)). It is important to note that the time and message complexity of COB depend only on the shortest source-destination distance and not on the overall network dimensions.

We also observe that both the random walk and the random waypoint mobility models result in

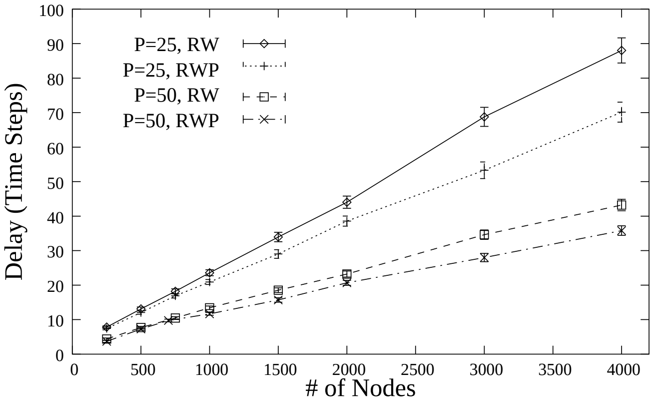

Fig. 5. Diameter scaling: Delay for a route discovery as a function of number of nodes N with proportional ($R \sim N$) network area of $R \times R$ m^2 for different (fixed) short-haul transmission ranges P m and mobility patterns (random walk (RW) and random waypoint (RWP))

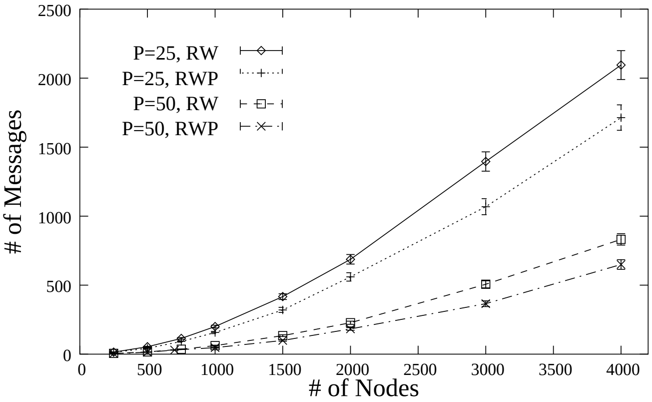

Fig. 6. Diameter scaling: Number of message transmissions for a route discovery as a function of number of nodes N with proportional ($R \sim N$) network area of $R \times R$ m^2 for different (fixed) short-haul transmission ranges P m and mobility patterns (random walk (RW) and random waypoint (RWP))

the same underlying asymptotic trends in the hop distance, delay, and message complexity. This is to be expected from our analysis of the COB algorithm which is general in that it does not assume any specific mobility model, ensuring that our theoretical results hold for any mobility behavior. The somewhat lower hop distance, delay, and number of messages for the random waypoint model observed in the plots are due to the slight tendency for the nodes to more densely populate the center of the network area with the random waypoint model, resulting in somewhat lower constants in the asymptotic scaling behavior.

B. Comparison with DSR

In this section, we compare the network layer performance of the COB route discovery process with the well-known Dynamic Source Routing (DSR) algorithm [50, 51]. We consider the node density scaling scenario where N nodes are uniformly distributed on an area of R = 500 m by R = 500 m. Nodes are freely moving in the area according to the random way point model with a randomly distributed speed in the range from 10 – 20 m/s and a pause time of 30 seconds. We consider the two transmission ranges P = 50 m and P = 100 m in DSR, which we consider to correspond to the short haul transmission ranges in COB. We consider the two performance metrics normalized routing load, which we define as the number of packets (messages) transmitted per data

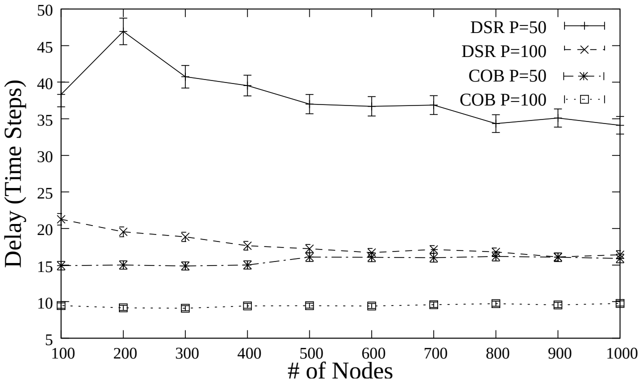

Fig. 7. Network layer comparison of COB and DSR: Delay for packet delivery as a function of number of nodes N for different (fixed) transmission ranges P (in m); pause time = 30 s, network area 500 $\times$ 500 m^2, fixed

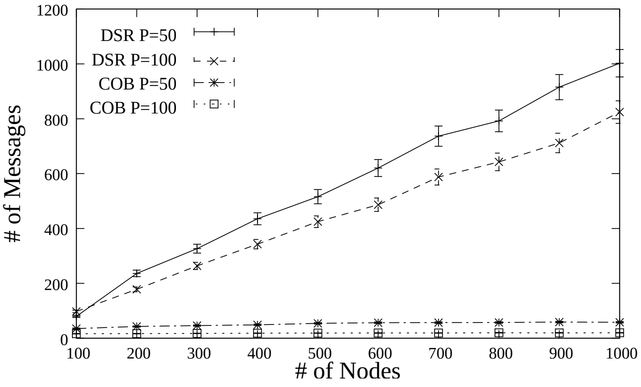

Fig. 8. Network layer comparison of COB and DSR: Normalized routing load as a function of number of nodes N for different (fixed) transmission ranges P (in m); pause time = 30 s, network area 500 $\times$ 500 m^2, fixed.

packet delivered to the destination, and mean delay, which we define as the number of time steps required to deliver a data packet from source to destination.

We observe from Figs. 7 and 8 that, as already observed above, the delay and the normalized routing load (message complexity) do not increase with the number of nodes in COB. On the other hand, we observe that the delay in DSR very slightly decreases as the number of nodes increases. This effect is caused by the slightly more crooked routes around uncovered areas with a small number of nodes, which diminishes with increasing network density. Note that DSR has a small delay for 100 nodes with a transmission range of P = 50 m. This is caused by the loss of some data packets due to the network not always being fully connected with only 100 nodes, whereby we count the delay only for successfully delivered packets. In other words, the part of the network that is connected and allows for successful delivery tends to have a somewhat smaller diameter. When comparing the delay values of COB and DSR it is important to keep in mind that in the considered setting the transmission range in DSR corresponds to the short-haul transmission range between a regular node and its cluster leader in COB. The long-haul transmissions in COB between adjacent cluster leaders have three times the range of the transmissions in DSR. Hence it may take three times as long with DSR to traverse the same distance as traversed with a long-haul transmission in COB. As observed in Fig. 7, the delay in time step units with DSR is between two and three

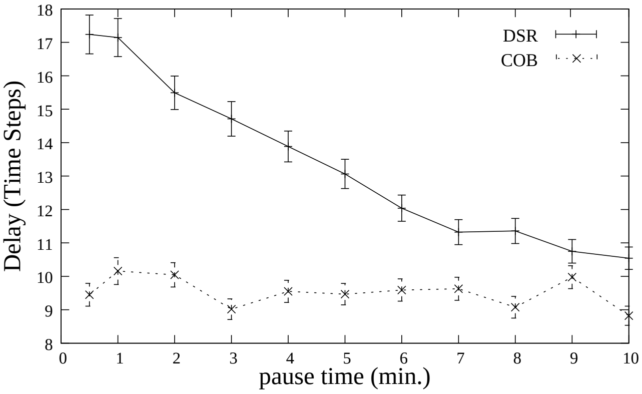

Fig. 9. Network layer comparison of COB and DSR: Delay for packet delivery as a function of pause time; transmission range P = 100 m, number of nodes N = 500, network area 500 × 500 m^2, fixed

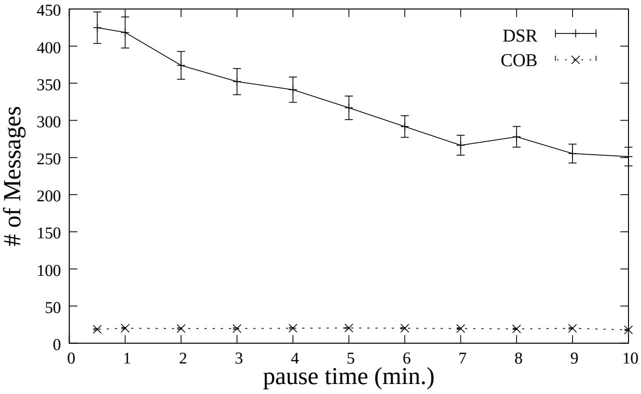

Fig. 10. Network layer comparison of COB and DSR: Normalized routing load as a function of pause time; transmission range P = 100 m, number of nodes N = 500, network area 500 × 500 m^2, fixed

times the delay with COB. This indicates that DSR may approximately achieve the same delay performance as a version of COB that uses short-haul transmissions via gateway nodes between adjacent cluster leaders.

We observe from Fig. 8 that the normalized routing load increases approximately linearly with the node density with DSR, which is caused by the increasing number of nodes within the transmission range of each given node, resulting in a larger total number of transmitted messages in the route discovery process. With increasing transmission range P, the normalized routing load decreases with both routing approaches, which is primarily due to the shorter hop distance between source and destination with the larger transmission range. Considering the scaling behavior of the number of transmitted messages with DSR in comparison with COB reveals the benefit of the constant density clustering used in COB, and indicates that it may be worthwhile to formally examine constant density clustering in the context of DSR (initial simulation explorations of clustering in conjunction with DSR are reported in [52, 53]).

Figs. 9 and 10 show the mean delay and normalized routing load as a function of the pause time in the mobility model for $N = 500$ nodes and a transmission range of $P = 100$ m. Since COB discovers a route from scratch whenever a node has data to send, the mean delay and the routing load are essentially independent of the pause time. With DSR, on the other hand, the delay and the routing load decrease as the pause time increases. This is because the cached routing table in DSR becomes stale less frequently with increasing pause time. For large pause times the delay with DSR (which uses only short-haul transmissions) approaches the delay with COB (which uses short-haul transmissions between regular source/destination node and cluster leader, and long-haul transmissions between cluster leaders), indicating that DSR may in fact give lower delays than a version of COB which uses only short-haul transmissions (via up to two gateway nodes between adjacent cluster leaders). This indicates that employing DSR's route caching mechanism on top of COB may result in an overall improved routing performance (at the expense of higher storage complexity in the nodes, which would need to be formally examined in future work).

C. Performance of COB over 802.11 MAC Layer

To assess the interactions of COB with the MAC layer and the performance with cross-traffic we conducted simulations with a model of the 802.11 MAC layer that includes cross-traffic, packet collisions, and movement during the route discovery phase. In these simulations with the MAC layer we consider a short-haul transmission range of $P = 50$ m, and corresponding long-haul transmission range of 150 m. We consider the delivery of 512 Byte data packets and set the transmission rate to 2 Mbps. Data packets are generated at each node according to an independent Poisson process with a rate that ensures that all cluster leaders, each of which works on one RREQ broadcast at a time, are always backlogged. Throughout, the nodes move according to the random waypoint mobility model with 1 m/s. Following [33] we assumed different frequency bands for the intra-cluster communication inside the individual clusters and the inter-cluster communication among adjacent cluster leaders. Our simulation model considers the distributed coordination function (DCF) of 802.11 which employs carrier sense multiple access with collision avoidance (CSMA/CA). We did not employ request-to-send/clear-to-send (RTS/CTS) reservations for the RREQ packets to avoid

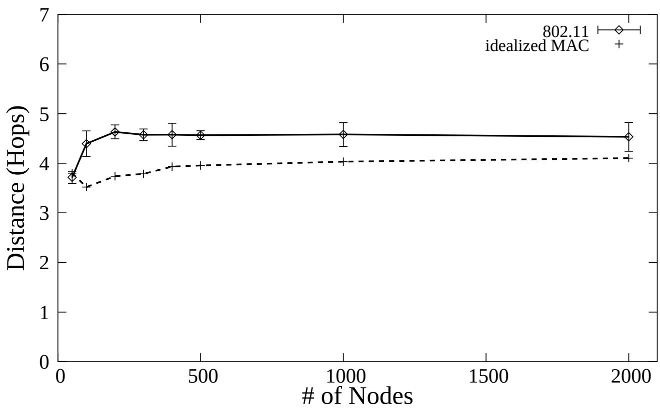

Fig. 11. Node density scaling: Length of discovered route as a function of number of nodes N with 802.11 and idealized MAC layers; network area 500 × 500 m^2, short-haul transmission range P = 50 m, fixed

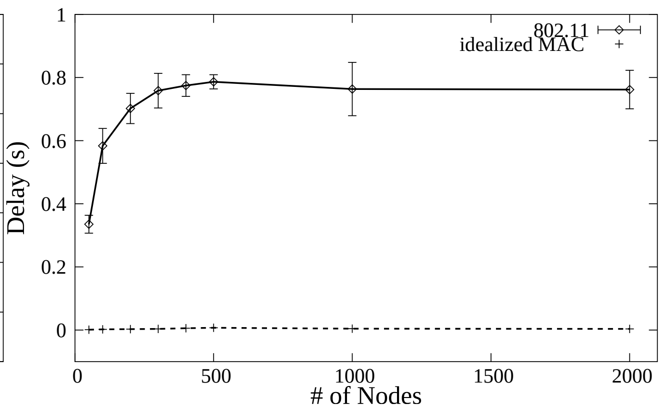

Fig. 12. Node density scaling: Delay for a packet delivery as a function of number of nodes N with 802.11 and idealized MAC layers; network area 500 × 500 m^2, short-haul transmission range P = 50 m, fixed

the reservation overhead for these short packets. If the doubling radius flooding passed an estimated maximum network diameter before reception of an ACK, the cluster leader moved on to the next queued request.

We considered the same two scaling scenarios as in Section V-A, and report statistics on the number of hops in the discovered route, the delay for the packet delivery, and the throughout (number of successfully delivered packets in 100 seconds of simulated network operation). Several statistically independent network operation periods were simulated and the obtained 95% confidence intervals are displayed in the plots.

The first series of results in Figs. 11, 12, and 13 for the node density scaling scenario indicate, in general, similar trends for the 802.11 MAC layer results and the idealized MAC layer results (which correspond to the network layer performance results from Section V-A). We observe from Fig. 11 that on average the discovered path lengths are somewhat longer with the 802.11 MAC layer. This effect is due to the collisions which may prevent routings from discovering the shortest path first. The overall trend, however, is for the discovered path length to remain relatively constant in both scenarios as the node density increases.

We observe from Fig. 12 that the packet delivery delay initially increases with the node density

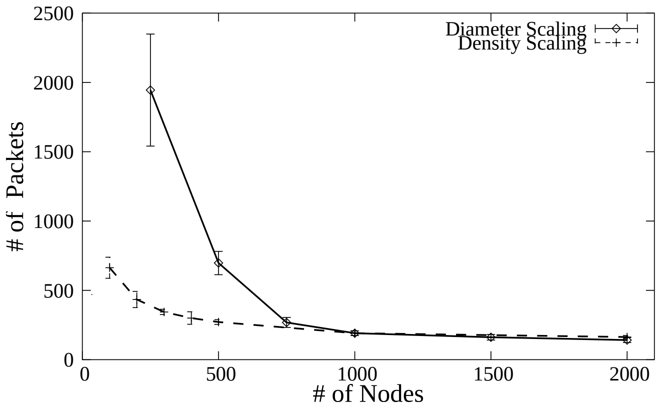

Fig. 13. Network density and diameter scaling: Throughput of successfully delivered data packets in 100 s with 802.11 MAC layer.

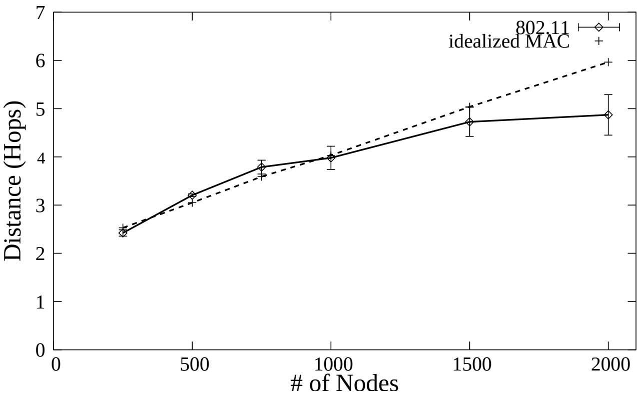

Fig. 14. Diameter scaling: Length of discovered route as a function of number of nodes N with proportional ($R \sim N$) network area of $R \times R$ (in m^2) with 802.11 and idealized MAC layer

and then flattens out for a high node density. This behavior resembles the behavior of the delays with the idealized MAC layer, which are constant when scaling up the node density, as shown in Fig 2. The absolute values of the 802.11 MAC layer delays are significantly larger than the delays with the idealized MAC layer. Higher density networks require our 802.11 MAC layer to spend more time re-broadcasting to successfully complete each one-hop transmission. Importantly the scaling behavior of the delay with the 802.11 MAC layer for increasing node density indicates that Theorem 2(a) still holds in that the delay (time complexity) is constant with respect to increasing node density. However, a more efficient MAC layer could most likely reduce the absolute values of the delays.

Fig. 13 gives the throughput in number of successfully delivered packets in 100 seconds of network operation. We observe an initial decrease in the throughout and then flattening out for high node densities. This behavior is primarily due to the increasing number of adjacent cluster leaders with increasing node density. More specifically, as the node density increases so does initially the number of adjacent cluster leaders. More adjacent cluster leaders result in more collisions in the long-haul transmissions, which in turn result in decreased throughput as seen in Fig. 13 (and increased delay as seen in Fig. 12). As noted in Section III, the LCC cluster algorithm ensures

TABLE I

DIAMETER SCALING: DELAY FOR A PACKET DELIVERY AS A FUNCTION OF NUMBER OF NODES N WITH PROPORTIONAL ($R \sim N$) NETWORK AREA OF $R \times R$ (IN M^2) WITH 802.11 AND IDEALIZED MAC LAYER

N	802.11 MAC (ms)	Idealized MAC (ms)
250	48.5	1.08
500	312.6	2.07
750	569.9	3.05
1000	763.5	3.88
1500	1193.7	5.89
2000	1636.3	7.49

that there are no more than 49 adjacent cluster leaders irrespective of the node density. As the node density increases further, the number of adjacent leaders approaches a maximum value and correspondingly the throughput approaches a constant level, as observed in Fig. 13.

Next, we examine the diameter scaling scenario, in which we jointly scale up the network area and number of nodes to achieve an approximately linear increase of the shortest hop distance Δ. We observe from Fig. 14 that with both MAC models the lengths of the discovered routes increase approximately linearly, tracking the increase of the shortest hop distance Δ. We observe from Table I a tendency toward linear scaling of the delay with both models, whereby the absolute delay values are orders of magnitude larger with the 802.11 MAC layer.

The throughout plotted in Fig. 13 exhibits again the initial drop, which is primarily due to the increasing hop distance (see Fig. 14), but then stabilizes even for large networks and correspondingly large hop distances. For the smallest simulated networks, the overlay network is quite compact, and hop distances often include only one or two long-haul transmissions. This vastly reduces congestion due to cross traffic, allowing for higher throughput. For larger networks, the overlay network of cluster leaders grows proportionally with the increasing number of nodes and network area in the diameter scaling scenario, allowing for relatively stable throughput levels even with increasing hop distance.

VI. CONCLUSIONS

We have developed and formally analyzed the Cluster Overlay Broadcast (COB) routing algorithm for MANETs. COB runs on top a cluster cover of the network with a constant density of cluster leaders, which we have proven can be maintained by the Least Cluster Change (LCC) algorithm. COB discovers routes with a doubling radius broadcast on the overlay network of cluster leaders. We note that the underlying mechanisms (routing on top of cluster cover, doubling radius technique) have also been examined in the context of the AODV routing protocol through simulations, see for instance [24, 32]. We have formally shown that by exploiting the constant density of the network of cluster leaders and the doubling radius technique, COB has a time complexity that is linear in the shortest source-destination hop distance and a message complexity that is quadratic in the shortest source-destination distance. Importantly, we have also shown that COB adapts optimally to the mobility of the nodes and has constant storage complexity in the nodes. Our theoretical results complement the existing simulation studies on MANET routing mechanisms and provide insight into the fundamental underpinnings of the performance of the routing mechanisms employed in COB and other protocols, such as AODV. To the best of our knowledge, COB is the first MANET routing algorithm that has been formally shown (i) to adapt optimally to the node mobility, and (ii) to have time and message complexities that are polynomial in the source-destination node distance and independent of the overall network size (total number of nodes, total diameter of network).

Our simulation results demonstrate that COB incurs essentially a constant delay as the number of nodes in a fixed network area (network density) is scaled up both when considering only the network layer as well as the network layer combined with an elementary 802.11 MAC layer. Also, with increasing network density, the number of message transmissions for a route discovery increases only until the network area is fully covered with clusters and then remains constant for further increasing node density. Our simulation results have also demonstrated that the delay scales linearly with the source-destination distance, and have indicated that the number of messages scales quadratically with the source-destination distance.

There are several broad areas for exciting future work on MANET routing and extensions to COB. One area is to integrate MANET routing in general, and the COB algorithm in particular, with higher layer notions of network services, such as service discovery or location and context aware services. In this context it is very interesting to examine the integration of routing with content distribution mechanisms, e.g., for providing multimedia services to MANET nodes. Another area is to develop cross-layer designs that integrate several network layers, possibly ranging from the application layer, including the network layer, and reaching down to the medium access and physical layers. Exploiting the specific characteristics of the wireless medium access and physical layers, similar to the approaches in [54–57], appears especially promising in these cross-layer designs. Throughout, we believe it is vital to pay close attention to and formally understand the scaling behaviors of the MANET protocols.

ACKNOWLEDGEMENT

We are grateful to the careful consideration of an earlier version of this manuscript by the three anonymous reviewers and their thoughtful comments on it, which helped us to significantly improve the quality of this article.

REFERENCES

[1] R. Wattenhofer, “Ad-hoc and sensor networks: worst-case vs. average-case,” in *Proceedings of International Zurich Seminar on Communications*, 2004, pp. 156–159.

[2] M. Abolhasan, T. Wysocki, and E. Dutkiewicz, “A review of routing protocols for mobile ad hoc networks,” *Ad Hoc Networks*, vol. 2, no. 1, pp. 1–22, Jan. 2004.

[3] X. Hong, K. Xu, and M. Gerla, “Scalable routing protocols for mobile ad hoc networks,” *IEEE Network*, pp. 11–21, July/August 2002.

[4] C. A. Santivanez, B. McDonald, I. Stavrakakis, and R. Ramanathan, “On the scalability of ad hoc routing protocols,” in *Proc. of IEEE Infocom*, June 2002, pp. 1688–1697.

[5] A. Iwata, C.-C. Chiang, G. Pei, M. Gerla, and T.-W. Chen, “Scalable routing strategies for ad hoc wireless networks,” *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 8, pp. 1369–1379, Aug. 1999.

[6] R. Rajamaran, “Topology control and routing in ad hoc networks: A survey,” *SIGACT News*, vol. 33, pp. 60–73, 2002.

[7] B. Karp and H. T. Kung, “GPSR: Greedy perimeter stateless routing for wireless networks,” in *Proceedings of ACM MobiCom*, Boston, MA, Aug. 2000, pp. 243–254.

[8] Y.-B. Ko and N. H. Vaidya, “Location-aided routing (LAR) in mobile ad hoc networks,” *Wirel. Netw.*, vol. 6, no. 4, pp. 307–321, 2000.

[9] S.-C. M. Woo and S. Singh, “Scalable routing protocol for ad hoc networks,” *Wirel. Netw.*, vol. 7, no. 5, pp. 513–529, 2001.

[10] J. Broch, D. Maltz, D. Johnson, Y.-C. Hu, and J. Jetcheva, "A performance comparison of multi-hop wireless ad hoc network routing protocols," in *Proceedings of the ACM Mobicom*, Dallas, TX, Oct. 1998, pp. 85–97.

[11] A. Boukerche, "Performance evaluation of routing protocols for ad hoc wireless networks," *Mobile Networks and Applications*, vol. 9, pp. 333–342, 2004.

[12] I. Chlamtac, M. Conti, and J. J. N. Liu, "Mobile ad hoc networking: imperatives and challenges," *Ad Hoc Networks*, vol. 1, no. 1, pp. 13–64, July 2003.

[13] W. Choi and S. K. Das, "Design and performance analysis of a proxy-based indirect routing scheme in ad hoc wireless networks," *Mob. Netw. Appl.*, vol. 8, no. 5, pp. 499–515, 2003.

[14] S. R. Das, R. Castaeda, and J. Yan, "Simulation-based performance evaluation of routing protocols for mobile ad hoc networks," *Mob. Netw. Appl.*, vol. 5, no. 3, pp. 179–189, 2000.

[15] J. J. Garcia-Luna-Aceves, M. Mosko, and C. Perkins, "Efficient on-demand loop-free routing in ad hoc networks," in *Proceedings of ACM Symposium on Principles of Distributed Computing (PODC)*, Boston, MA, July 2003.

[16] S.-J. Lee, J. Hsu, R. Hayashida, M. Gerla, and R. Bagrodia, "Selecting a routing strategy for your ad hoc network," *Computer Communications*, vol. 26, no. 7, pp. 723–733, May 2003.

[17] E. M. Belding-Royer and C. E. Perkins, "Evolution and future directions of the ad hoc on-demand distance-vector routing protocol," *Ad Hoc Networks*, vol. 1, no. 1, pp. 125–150, July 2003.

[18] S. J. Philip, J. Ghosh, S. Khedekar, and C. Qiao, "Scalability analysis of location management protocols for mobile ad hoc networks," in *Proceedings of IEEE Wireless Communications and Networking Conference*, Mar. 2004, pp. 183–188.

[19] A. Sankar and Z. Liu, "Maximum lifetime routing in wireless ad-hoc networks," in *Proceedings of IEEE Infocom*, Hong Kong, Mar. 2004.

[20] A. Srinivas and E. Modiano, "Minimum energy disjoint path routing in wireless ad-hoc networks," in *Proceedings of ACM MobiCom*, 2003, pp. 122–133.

[21] J. Sucec and I. Marsic, "Clustering overhead for hierarchical routing in mobile ad hoc networks," in *Proc. of IEEE Infocom*, June 2002, pp. 1698–1706.

[22] H. Wu and A. Abouzeid, "Cluster-based routing overhead in networks with unreliable nodes," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2004.

[23] G. Pei, M. Gerla, X. Hong, and C. Chiang, "A wireless hierarchical routing protocol with group mobility," in *Proceedings of IEEE Wireless Communications and Networking*, New Orleans, LA, Sept. 1999, pp. 1538–1542.

[24] E. M. Belding-Royer, "Multi-level hierarchies for scalable ad hoc routing," *Wirel. Netw.*, vol. 9, no. 5, pp. 461–478, 2003.

[25] C.-C. Chiang and M. Gerla, "Routing and multicast in multihop, mobile wireless networks," in *Proc. IEEE Int. Conference on Universal Personal Communications*, San Diego, CA, Oct. 1997, pp. 546–551.

[26] Z. Haas and M. R. Pearlman, "The performance of query control schemes for the zone routing protocol," *IEEE/ACM Transactions on Networking*, vol. 9, no. 4, pp. 427–438, 2001.

[27] M. Jiang, J. Ji, and Y. C. Tay, "Cluster based routing protocol, internet draft, work in progress," 1999.

[28] P. Krishna, N. H. Vaidya, M. Chatterjee, and D. K. Pradhan, "A cluster-based approach for routing in dynamic networks," *SIGCOMM Comput. Commun. Rev.*, vol. 27, no. 2, pp. 49–64, 1997.

[29] R. Ramanathan and M. Steestrup, "Hierarchically organized multihop mobile wireless networks for quality of service support," *Mobile Networks and Applications*, vol. 3, no. 1, pp. 101–119, 1999.

[30] P. Sinha, R. Sivakumar, and V. Bharghavan, "CEDAR: a core-extraction distributed ad hoc routing algorithm," in *Proc. of IEEE Infocom*, New York, NY, Mar. 1999, pp. 202–209.

[31] Y. Yi and M. Gerla, "Scalable AODV with efficient flooding based on on-demand clustering," *ACM Mobile Computing and Communications Review*, vol. 6, no. 3, pp. 98–99, 2002.

[32] S.-J. Lee, E. M. Belding-Royer, and C. E. Perkins, "Scalability study of the ad hoc on-demand distance vector routing protocol," *Int. J. Netw. Manag.*, vol. 13, no. 2, pp. 97–114, 2003.

[33] K. Xu and M. Gerla, "A heterogeneous routing protocol based on a new stable clustering scheme," in *Proc. of IEEE Milcom*, Oct. 2002, pp. 838–843.

[34] R. Castaneda and S. Das, "Query localization techniques for on-demand routing protocols in ad hoc networks," in *Proceedings of ACM MobiHoc*, Seattle, WA, 1999, pp. 186–194.

[35] H. Dubois-Ferriere, M. Grossglauser, and M. Vetterli, "Age matters: Efficient route discovery in mobile ad hoc networks using encounter ages," in *Proceedings of ACM MobiHoc*, Annapolis, MD, June 2003, pp. 257–266.

[36] J. Eriksson, M. Faloutsos, and S. Krishnamurthy, "Scalable ad hoc routing: The case for dynamic addressing," in *Proceedings of IEEE Infocom 2004*, Hong Kong, Mar. 2004.

[37] C. Gui and P. Mohapatra, "SHORT: self-healing and optimizing routing techniques for mobile ad hoc networks," in *Proceedings of ACM MobiHoc*, 2003, pp. 279–290.

[38] K. Alzoubi, P.-J. Wan, and O. Frieder, "New Distributed Algorithm for Connected Dominating Set in Wireless Ad Hoc Networks," in *Proceedings of the Thirty-Fourth Annual Hawaii International Conference on System Science (HICSS-35)*, 2002.

[39] D. Baker, A. Ephremides, and J. A. Flynn, "The design and simulation of a mobile radio network with distributed control," *IEEE Journal on Selected Areas in Communications*, vol. 2, no. 1, pp. 226–237, Jan. 1984.

[40] S. Basagni, "Distributed and mobility-adaptive clustering for multimedia support in multi-hop wireless networks," in *Proc. IEEE Vehicular Tech. Conf.*, 1999, pp. 19–22.

[41] G. Calinescu, I. I. Mandoiu, P.-J. Wan, and A. Z. Zelikovsky, "Selecting forwarding neighbors in wireless ad hoc networks," *Mob. Netw. Appl.*, vol. 9, no. 2, pp. 101–111, 2004.

[42] M. Chatterjee, S. K. Das, and D. Turgut, "WCA: A weighted clustering algorithm for mobile ad hoc networks," *Cluster Computing*, vol. 5, no. 2, pp. 193–204, 2002.

[43] A. B. McDonald and T. F. Znati, "A mobility-based framework for adaptive clustering in wireless ad hoc networks," *IEEE Journal on Selected Areas on Communications*, vol. 17, no. 8, pp. 1466–1487, Aug. 1999.

[44] X.-Y. Li, Y. Wang, P.-J. Wan, and O. Frieder, “Localized low weight graph and its applications in wireless ad hoc networks,” in *Proceedings of IEEE Infocom 2004*, Hong Kong, Mar. 2004.

[45] O. Younis and S. Fahmy, “Distributed clustering in ad-hoc sensor networks: A hybrid, energy-efficient approach,” in *Proceedings of IEEE Infocom 2004*, Hong Kong, Mar. 2004.

[46] C.-C. Chiang, H.-K. Wu, W. Liu, and M. Gerla, “Routing in clustered multihop, mobile wireless networks with fading channel,” in *Proc. IEEE Singapore Int. Conf. on Networks*, 1997, pp. 197–211.

[47] H. Huang, A. W. Richa, and M. Segal, “Approximation algorithms for the mobile piercing set problem with applications to clustering in ad hoc networks,” *Mobile Networks and Applications (MONET)*, vol. 9, no. 2, pp. 151–161, Apr. 2004.

[48] A. D. Amis and R. Prakash, “Load-balancing clusters in wireless ad hoc networks,” in *Proceedings of IEEE Symposium on Application-Specific Systems and Software Engineering Technology*, Mar. 2000, pp. 25–32.

[49] A. Safwat, H. Hassanein, and H. Mouftah, “Power-aware fair infrastructure formation for wireless mobile ad hoc communications,” in *Proceedings of IEEE Globecom*, Nov. 2001, pp. 2832–2836.

[50] D. B. Johnson, D. A. Malts, and J. Broch, “DSR: The dynamic source routing protocol for multi-hop wireless ad hoc networks,” in *Ad Hoc Networking, Chapter 5*, C. E. Perkins, Ed. Addison–Wesley, 2001, pp. 139–172.

[51] D. B. Johnson, D. A. Malts, and Y.-C. Hu, “The dynamic source routing protocol for mobile ad hoc networks (DSR),” IETF Internet Draft, Tech. Rep., July 2004, draft-ietf-manet-dsr-10.txt.

[52] M. Soyturk, E. Cayirci, and A. E. Harmanci, “Virtual cell layout based dynamic source routing algorithm for the mobile subsystem of the next generation tactical communications systems,” in *Proceedings of IEEE Milcom*, Oct. 2002, pp. 541–545.

[53] Y. Yi, M. Gerla, and T. J. Kwon, “The selective intermediate nodes scheme for ad hoc on-demand routing protocols,” in *Proceedings of IEEE ICC*, May 2002, pp. 3191–3195.

[54] A. Muqattash and M. M. Krunz, “A distributed transmission power control protocol for mobile ad hoc networks,” *IEEE Transactions on Mobile Computing*, vol. 3, no. 2, pp. 113–128, April-June 2004.

[55] B. Sadeghi, V. Kanodia, A. Sabharwal, and E. Knightly, “Opportunistic media access for multirate ad hoc networks,” in *Proceedings of ACM/IEEE Mobicom*, Atlanta, GA, Sept. 2002.

[56] J. A. Stine and G. de Veciana, “A paradigm for quality-of-service in wireless ad hoc networks using synchronous signaling and node states,” *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 7, pp. 1301–1321, Sept. 2004.

[57] X. Yang and N. Vaidya, “On physical carrier sensing in wireless ad hoc networks,” in *Proc. of IEEE Infocom*, Miami, FL, Mar. 2005.