# Distributed Low-Cost Backbone Formation for Wireless Ad Hoc Networks

Yu Wang
Dept. of Computer Science
University of North Carolina
Charlotte, NC 28223, USA
ywang32@uncc.edu

Weizhao Wang
Dept. of Computer Science
Illinois Institute of Technology
Chicago, IL 60616, USA
wangwei4@iit.edu

Xiang-Yang Li *
Dept. of Computer Science
Illinois Institute of Technology
Chicago, IL 60616, USA
xli@cs.iit.edu

## ABSTRACT

Backbone has been used extensively in various aspects (*e.g.*, routing, route maintenance, broadcast, scheduling) for wireless networks. Previous methods are mostly designed to minimize the backbone size. However, in many applications, it is desirable to construct a backbone with small *cost* when each wireless node has a cost of being in the backbone. In this paper, we first show that previous methods specifically designed to minimize the backbone size may produce a backbone with a large cost. We then propose an efficient distributed method to construct a weighted sparse backbone with low cost. We prove that the total cost of the constructed backbone is within a small constant factor of the optimum for homogeneous networks when either the nodes' costs are smooth or the network maximum node degree is bounded. We also show that with a small modification the constructed backbone is efficient for unicast: the total cost (or hop) of the least cost (or hop) path connecting any two nodes using backbone is no more than 3 (or 4) times of the least cost (or hop) path in the original communication graph. As a side product, we give an efficient overlay based multicast structure whose total cost is no more than 10 times of the minimum when the network is modeled by UDG. Our theoretical results are corroborated by our simulation studies.

## Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Network topology, Wireless communication; G.2.2 [**Graph Theory**]: Network problems, Graph algorithms.

## General Terms

Algorithms, Design, Performance, Theory.

## Keywords

Connected dominating set, clustering, weighted, localized algorithm, wireless ad hoc/sensor networks.

*The work of the author is partially supported by NSF CCR-0311174.


*MobiHoc'05*, May 25–27, 2005, Urbana-Champaign, Illinois, USA.


## 1. INTRODUCTION

Wireless networks draw lots of attentions in recent years due to its potential applications in various areas. Many routing protocols have been proposed for wireless ad hoc networks recently. The simplest routing method is to flood the message, which not only wastes the rare resources of wireless nodes, but also diminishes the throughput of the network. One way to avoid flooding is to let each node communicate with only a selected subset of its neighbors , or to use a hierarchical structure like Internet, *e.g.*, connected dominating set (CDS) based routing [15, 30, 35, 41].

Efficient distributed algorithms for constructing connected dominating sets in ad hoc wireless networks were well studied [3, 4, 5, 7, 15, 37, 41]. The notion of cluster organization has been used for wireless ad hoc networks since their early appearance. Baker *et al.* [4, 5] introduced a fully distributed linked cluster architecture mainly for hierarchical routing and demonstrated its adaptivity to the network connectivity changes. The notion of the cluster has been revisited by Gerla *et al.* [18, 31] for multimedia communications with the emphasis on the allocation of resources, namely, bandwidth and channel, to support the multimedia traffic in an ad hoc environment. In [17], Gao, *et al.* proposed a randomized algorithm for maintaining the discrete mobile centers, *i.e.*, dominating sets. They showed that it approximates *minimum dominating set* (MDS) within $O(1)$ with very high probability. Recently, Alzoubi *et al.* [3, 39] proposed a method to approximate *minimum connected dominating set* (MCDS) within 8 whose message complexity is $O(n \log n)$ and time complexity is $O(n)$ for wireless networks modeled by unit disk graphs. Alzoubi, *et al.* [2] continued to propose a localized method approximating the MCDS within a constant time using a linear number of messages. Existing clustering methods first choose some nodes to act as coordinators of the clustering process, *i.e.*, clusterhead. Then a cluster is formed by associating the clusterhead with some (or all) of its neighbors. Previous methods differ on the criterion for the selection of the clusterhead, which is either based on the lowest (or highest) ID among all unassigned nodes [4, 31], or based on the maximum node degree [18], or based on some generic weight [7] (the node with the largest weight will be chosen as clusterhead). In [1], the authors proposed to build the local Delaunay graph on top of an approximated MCDS for efficient routing. Recently, Kuhn and Wattenhofer [27] proposed a new distributed MDS approximation algorithm based on linear programming (LP) relaxation techniques. For an arbitrary parameter $k$ and maximum degree $\Delta$, their algorithm computes a dominating set of expected size $O(k\Delta^{2/k} \log \Delta |MDS|)$ in $O(k^2)$ rounds where each node has to send $O(k^2\Delta)$ messages of size $O(\log \Delta)$. Moreover, the authors further gave the time lower bounds for the distributed approximation of MDS in [25]. In [26], the authors

show how to compute a good dominating set in a harsh model in which there is no underlying MAC layer, asynchronous wake-up, scarce knowledge about the network topology.

All of above methods try to minimize the number of clusterheads, *i.e.*, the number of nodes in the backbone. However, in many applications of wireless ad hoc networks, minimizing the size of the backbone is not sufficient. For example, different wireless nodes may have different costs for serving as a clusterhead, due to device differences, power capacities, and information loads to be processed. Therefore, in the remaining of the paper, for the succinctness of our presentation, we assume that each wireless node has a *generic cost* (or *weight*). The cost may also represent the *fitness* or *priority* of each node to be a clusterhead. The lower cost means the higher priority. In practice, the cost could represent the power consumption rate of this node if a backbone with small power consumption is needed; the robustness of this node if a fault-tolerant backbone is needed; or a function of its security level if a secure backbone is needed. We study how to construct a sparse backbone efficiently for a set of weighted wireless nodes such that the total cost of the backbone is minimized and there is a cost (or hops) *efficient* route connecting every pair of wireless nodes via the constructed network backbone. Here a route is cost (or hops resp.) *efficient* if its cost (or hops resp.) is no more than a constant factor of the minimum cost (or hops resp.) needed to connect the source and the destination in the original communication graph when all possible physical communication links are considered.

Recently, many proposed clustering algorithms [6, 7, 8, 10, 12, 13, 21, 24, 32, 34, 36, 38] also considered different weights as a *priority criterion* to decide whether a node will be a clusterhead. Notice, the ultimate goal of the majority protocols is still to minimize the size of the cluster (or backbone), not the total weight of the cluster (or backbone). For example, methods in [7, 10, 34] considered the stability or mobility of each node as the weight. They preferred the node with high stability and low mobility to be the clusterhead. However, the definitions of stability or the evaluation methods used are different. In [24], authors also combined the stability with the degree of each node as the weight. The higher priority is given to relatively stable and high degree nodes. Methods in [21, 36] considered clustering in heterogeneous sensor networks, where each node has different energy level. Most of them used the remaining energy or energy consumption rate as the weight. Both [6] and [32] considered two factors in the priority: available energy and the speed, though they used different equations to combine them. In [12], Chatterjee *et al.* considered a combined weight metric for their clustering algorithm, that takes into account several system parameters like the node-degree, transmission power, mobility and the battery power of the nodes. Similarly, Nocetti *et al.* [13] also combined these four facts to be the weights for their clustering method. A nice literature review of cluster methods can be found in [13]. In [9], Basagni *et al.* also showed the performance comparison of some proposed protocols for clustering and backbone formation. Most of these proposed weighted clustering algorithms applied the simple greedy algorithms where the nodes with highest priority (lowest cost) become clusterheads. For example, cluster method in [12] selects a node with the lowest cost among its unchosen neighbors to serve as a clusterhead. These greedy heuristics work well in practice, but as we will show in Section 2 that they may generate a backbone with a high cost compared with the optimum. Some of these methods [21, 36] are randomized algorithms, nodes become clusterheads randomly with a weighted election probability. In [38], Turgut proposed a genetic algorithm to optimize cluster processing. All of these cluster methods do not guarantee any approximation ratio of the weighed cluster (or backbone) compared with the optimum. Notice that, in [8], Basagni gave an algorithm to solve *maximal weighted independent set* in wireless networks, but here our solution for cluster is a distributed approximation algorithm for *maximum weighted independent set*, and *minimum connected dominating set* which are well-known NP-hard problems. Li *et al.* [28] presented a centralized approximation algorithm for weighted maximum independent set for some special graphs. Guha and Khuller [19] studied centralized algorithms for weighted minimum connected dominating set in general graphs, by combining a weighted set cover approximation algorithm and a node-weighted Steiner tree approximation algorithm they achieved approximation ratio $3 \ln n$. In [20], they further improved the approximation ratio to $1.35 \ln n$ which is the best known ratio. In addition, any approximation algorithm with ratio $\alpha$ for the unweighted (connected) dominating set problem automatically gives ratio $\alpha \cdot \delta$ for the weighted version. In particular, the known PTAS for dominating set in UDG [22] implies that weighted dominating set in UDG can be approximated with ratio $(1 + \epsilon) \cdot \delta$ for arbitrary $\epsilon > 0$.

In this paper, we propose a novel distributed method to generate weighted backbone with a good approximation ratio while using small communication cost. Our methods work not only for homogeneous networks, but also for heterogeneous networks. We prove that the total cost of the constructed backbone is within $\min(4\delta + 1, 18 \log(\Delta + 1)) + 10$ times of the optimum for homogeneous networks when all nodes have the same transmission range. Here $\delta$ is the maximum ratio of costs of two adjacent wireless nodes and $\Delta$ is the maximum node degree in the communication graph. Notice that the advantage of our backbone is that the total cost is small compared with the optimum when either the costs of wireless nodes are smooth, *i.e.*, two neighboring nodes' cost differ by a small constant factor, or the maximum node degree is low. The total number of messages of our method is $O(m)$ for any network composed of $n$ wireless devices and $m$ total pairs of nodes that can directly receive signals from each other. We also show that with a small modification the constructed backbone is efficient for unicast: the total cost (or hop) of the least cost (or hop) path connecting any two nodes using backbone is no more than 3 (or 4) times of the least cost (or hop) path in the original communication graph. This is significant since our backbone structure is much sparser than the original communication graph, which significantly reduces the cost of routing without losing much ground on the performance of unicast.

The rest of the paper is organized as follows. In Section 2, we provide preliminaries necessary for describing our new algorithms, and show the possible bad performances of several proposed methods. Section 3 presents our new weighted backbone formation algorithms, and Section 4 gives the theoretical performance analysis of the proposed algorithms. In Section 5, we discuss several possible network applications of our proposed weighted backbone formation algorithms. Section 6 presents the experimental results. We conclude our paper in Section 7 by pointing out some possible future research directions.

## 2. PRELIMINARIES AND RELATED WORKS

In this section, we first give some definitions and notations that will be used in our presentation later. We assume that all wireless nodes are given as a set $V$ of $n$ points in a two dimensional space. Each wireless node has an omni-directional antenna. This is attractive for a single transmission of a node can be received by all nodes within its vicinity. Each node has some computational power. We always assume that the nodes are almost-static in a reasonable period of time. A communication graph $G = (V, E)$ over a set $V$

of wireless nodes has an edge $uv$ between nodes $u$ and $v$ iff $u$ and $v$ can communicate directly with each other, *i.e.*, inside the transmission region of each other. Hereafter, we always assume that $G$ is a connected graph. Let $d_G(u)$ be the degree of node $u$ in a graph $G$ and $\Delta$ be the maximum node degree of all wireless nodes (i.e. $\Delta = \max_{u \in V} d_G(u)$). Notice that the average node degree is called *density* of the network. We assume that each wireless node $u$ has a cost $c(u)$ of being in the backbone. Here the cost $c(u)$ could be the value computed based on a combination of its remaining battery power, its mobility, its node degree in the communication graph, and so on. We will discuss several possible weight functions for different applications in Section 5 in detail. In general, smaller $c(u)$ means that the node is more suitable of being in the backbone. Let $\delta = \max_{i,j \in E} c(i)/c(j)$, where $E$ is the set of communication links in the wireless network $G$. We call $\delta$ the *cost smoothness* of the wireless networks. When $\delta$ is bounded by some small constant, we say the node costs are *smooth*.

When the transmission region of every wireless node is modeled by a unit disk centered at itself, the communication graph is often called a *unit disk graph*, denoted by $UDG(V)$, in which there is an edge between two nodes if and only if their distance is at most one. We also call such wireless networks as *homogeneous networks*.

We call all nodes within a constant $k$ hops of a node $u$ in the communication graph $G$ as the *$k$-local nodes* or *$k$-hop neighbors* of $u$, denoted by $N_k(u)$, which includes $u$ itself. The $k$-local graph of a node $u$, denoted by $G_k(u)$, is the induced graph of $G$ on $N_k(u)$, *i.e.*, $G_k(u)$ is defined on vertex set $N_k(u)$, and contains all edges in $G$ with both end-points in $N_k(u)$.

A subset of vertices in a graph $G$ is an *independent set* if for any pair of vertices, there is no edge between them. It is a *maximal independent set* if no more vertices can be added to it to generate a larger independent set. It is a *maximum independent set* (MIS) if no other independent set has more vertices. The independence number, denoted as $\alpha(G)$, of a graph $G$ is the size of the maximum independent set of $G$. The *$k$-local independence number*, denoted by $\alpha^{[k]}(G)$, is defined as $\alpha^{[k]}(G) = \max_{u \in V} \alpha(G_k(u))$. It is well-known that for a unit disk graph, $\alpha^{[1]}(UDG) \leq 5$ and $\alpha^{[2]}(UDG) \leq 18$.

A subset $S$ of $V$ is a *dominating set* if each node $u$ in $V$ is either in $S$ or is adjacent to some node $v$ in $S$. Nodes from $S$ are called dominators, while nodes not in $S$ are called dominatees. Clearly, any maximal independent set is a dominating set. A subset $C$ of $V$ is a *connected dominating set* (CDS) if $C$ is a dominating set and $C$ induces a connected subgraph. Consequently, the nodes in $C$ can communicate with each other without using nodes in $V - C$. A dominating set with minimum cardinality is called *minimum dominating set* (MDS). A connected dominating set with minimum cardinality is the *minimum connected dominating set* (MCDS).

In wireless ad hoc networks, assume that each node $u$ has a cost $c(u)$. Then a connected dominating set $C$ is called *weighted connected dominating set* (WCDS). A subset $C$ of $V$ is a *minimum weighted connected dominating set* (MWCDS) if $C$ is a WCDS with minimum total cost.

Several methods have been proposed in the literature to find a small dominating set for homogeneous networks. Most of them are based on greedy algorithms. Since, in this paper, we are interested in distributed methods, we will thus mainly discuss the priori distributed greedy methods here. If we insist on applying these distributed methods to approximate the minimum weighted dominating set, they may produce a backbone that is arbitrarily worse than the optimum. We will show by examples that three classical methods do not generate a dominating set whose cost is always comparable with ours in the worst case.

The first method to generate a dominating set is to generate a maximal independent set as follows [1, 12]. First, assume that all nodes are marked as WHITE originally, which represents that the node is not assigned any role yet. A node $u$ sends a message IamDominator to all its one-hop neighbors if it has the smallest cost (ID is often used if every node has a unit cost) among all its WHITE neighbors. Node $u$ also marks itself Dominator. When a node $v$ received a message IamDominator from its one-hop neighbors, node $v$ then marks itself Dominatee. Node $v$ then sends a message IamDominatee to all its one-hop neighbors. Clearly, the nodes marked with Dominator indeed form a dominating set.

We then show by example that the produced dominating set may be arbitrarily larger than the optimum solution. Although the instance illustrated here uses UDG as communication graph, it is not hard to extend this to general communication graph. See Figure 1 for an illustration. Assume that 3 wireless nodes $u$, $v$ and $w$ are distributed along a line with one unit interval. The nodes' costs of $u$, $v$, and $w$ are $\infty$, 1, and $1 - \epsilon$ respectively. The dominators selected by the first method are nodes $w$ and $u$, and the total cost of the solution is $\infty$. However, the optimal solution is formed by $v$ with a total cost 1. Our method presented later does produce a dominating set of total cost $2 - \epsilon$.

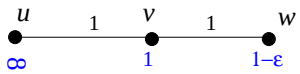



**Figure 1: An example why the first method fails to produce low cost weighted connected dominating set.**

The second method of constructing a dominating set [15, 16] is based on minimum weighted set cover [14]. The method can be described in a centralized way as follows: in each round, we select an unselected node $i$ with the minimum ratio $c(i)/d_i$, where $d_i$ is the number of nodes not covered by previously selected dominators. It is well-known that this centralized method produces a dominating set whose total cost is no more than $\log(\Delta + 1)$ times of the optimum, where $\Delta$ is the maximum original degree of all nodes. In [3], Alzoubi *et al.* gave an example (as in Figure 2) with a family of instances for which the size of the solution computed by the second method is larger than the optimum solution by a logarithm factor when all nodes have the same weight. Although the instance illustrated here uses UDG as communication graph, obviously, we can extend this to a general communication graph. In this example, all nodes have a unit weight. For the detail of this example, see [3]. Moreover, this method is expensive to implement in a distributed way. First, it is expensive to find the node $i$ with the minimum ratio $c(i)/d_i$ among all unchosen nodes. Second, it is also expensive to update $d_i$, which is the number of neighbors that are not covered by previously selected dominators. Our method described later will produce a dominating set whose size is no more than 5 times of the optimum for unit weighted UDG. More importantly, our method is a fully distributed method.

The third method to select the dominating set is proposed by Bao and Garcia-Luna-Aceves [6]. Unlike the previous two methods, this is a fully localized method and it can be executed in 2 rounds using synchronous communication model. A node decides to become a dominator if either one of the following two criteria are satisfied: 1) the node has the smallest cost in its one-hop neighborhood; 2) the node has the smallest cost in the one-hop neighborhood of one of its one-hop neighbors. We show by an example that the produced dominating set may be arbitrarily larger than the optimum solution. See Figure 3 for an illustration of an instance in UDG. Assume that $2n + 1$ wireless nodes are distributed as shown

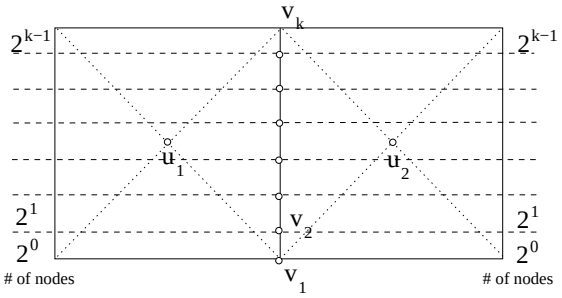



**Figure 2: An example [2] why the second method fails to produce low cost weighted connected dominating set.**

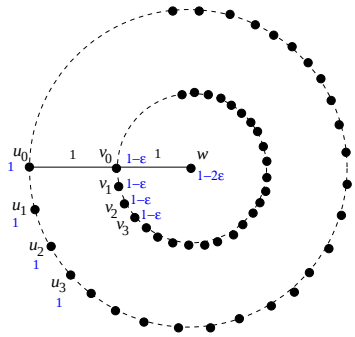



**Figure 3: An example why the third method fails to produce low cost weighted connected dominating set.**

in Figure 3. The nodes' costs of $u_i$, $v_i$, and $w$ are 1, $1-\epsilon$, and $1-2\epsilon$, respectively. The dominators selected by the third method are nodes $w$ and $v_i$ ($0 \leq i < n$), the total cost of the solution is $n(1-\epsilon)+1-2\epsilon$. However, the optimal solution formed by node $w$ and seven nodes from $u_i$ has total cost $8-2\epsilon$. It is easy to show that seven unit disks centered at 7 nodes among some $u_i$ can cover all $u_i$. Our method described later will produce an optimal dominating set in this special case.

# 3. EFFICIENT LOW-COST BACKBONE FORMATION ALGORITHMS

In this section, we will propose a distributed algorithm that can construct a low-cost backbone (weighted connected dominating set) for a wireless ad hoc network $G$ by assuming that each wireless node $u$ has a cost $c(u)$ of being on the backbone. We will prove that the total cost of the constructed backbone is no more than

$$\min(\alpha^{[2]}(G)\log(\Delta+1), (\alpha^{[1]}(G)-1)\delta+1)+2\alpha^{[1]}(G)$$

times of the optimum solution. Here $\Delta$ is the maximum degree of all wireless nodes, and $\delta = \max_{i,j\in E} c(i)/c(j)$, where $E$ is the set of communication links in the wireless network. Notice that, for homogeneous wireless networks modeled by UDG, it implies that the backbone produced by our method has a cost no more than $\min(18\log(\Delta+1), 4\delta+1)+10$ times of the optimum solution.

Here, we assume that each node knows the IDs and costs of all its 1-hop neighbors, which can be achieved by requiring each node to broadcast its ID and cost to its 1-hop neighbors initially. This protocol can be easily implemented using synchronous communications as did in [4, 5]. If the number of neighbors of each node is known a priori, then this protocol can also be implemented using asynchronous communications. Our method has the following two phases: the first phase (clustering phase) is to find a set of wireless nodes as the dominators[1] and the second phase is to find a set of nodes, called *connector*, to connect these dominators to form the final backbone of the network. Notice that these two phases could interleave in the actual construction method. We separate them just for the sake of easy presentations.

## 3.1 Finding Dominators

We then propose our method of constructing a dominating set whose total cost is comparable with the optimum solution. Our method first constructs a maximal independent set (MIS) using node weight as selection criterion. Then for each node $v$ in MIS, we run local greedy set cover method on *local neighborhood* $N_2(v)$ to find some nodes $GRDY_v$ to cover all one-hop neighbors of $v$. If $GRDY_v$ has a total cost smaller than $v$, then we use $GRDY_v$ to replace $v$, which will further reduce the cost of MIS. Our method works as follows.

**Algorithm 1** Construct Low-cost Dominating Set

1: First assume that all nodes are originally marked WHITE.
2: A node $u$ sends a message ItryDominator to all its one-hop neighbors if it has the smallest cost among all its WHITE neighbors. Node $u$ also marks itself PossibleDominator.
3: When a node $v$ received a message ItryDominator from its one-hop neighbors, node $v$ then marks itself Dominatee. Node $v$ then sends a message IamDominatee to all its one-hop neighbors.
4: When a node $w$ receives a message IamDominatee from its neighbor $v$, node $w$ removes node $v$ from its list of WHITE neighbors.
5: Each node $u$ marked with PossibleDominator collects the cost and ID of all of its two-hop neighbors $N_2(u)$.
6: Using the greedy method for minimum weighted set cover (like the second method), node $u$ selects a subset of its two hop-neighbors to cover *all* the one-hop neighbors (including $u$) of node $u$. If the cost of the selected subset, denoted by $GRDY_u$, is smaller than the cost of node $u$, then node $u$ sends a message YouAreDominator($w$) to each node $w$ in the selected subset. Otherwise, node $u$ just marks itself Dominator.
7: When a node $w$ received a message YouAreDominator($w$), node $w$ marks itself Dominator.

For the example illustrated by Figure 1, the MIS will be two nodes $w$ and $u$, whose cost is large. Node $u$ is PossibleDominator and thus performs the local set cover. Clearly $N_2(u)=\{u,v,w\}$ and $N_1(u)=\{u,v\}$. The local set cover will select $v$ to cover all nodes in $N_1(u)$ since $v$ covers both nodes in $N_1(u)$. Note that $c(v)<c(u)$, so node $u$ will let $v$ be a dominator. The other PossibleDominator $w$ will keep itself as a dominator since the local set cover gets worse solution than itself. The final dominating set is then $\{v,w\}$, which is close to optimum $\{v\}$.

## 3.2 Finding Connectors

The second step of weighted connected dominating set formation is to find some *connectors* (also called *gateways*) among all the dominatees to connect the dominators. Then the connectors and the dominators form a *connected dominating set* (or called backbone).

[1]We will interchange the terms cluster-head and dominator. The node that is not a cluster-head is also called *ordinary* node or *dominatee*. A node is called *white* node if its status is yet to be decided by the clustering algorithm. Initially, all nodes are white. The status of a node, after the clustering method finishes, could be *dominator* or *dominatee*.

Several methods [1, 2, 4, 5, 18] have been proposed in the literature to find the connectors. However, all of these methods only consider the unweighted scenario. We can show by examples that these methods generally do not produce a weighted connected dominating set with good approximation ratio.

Given a dominating set $S$, let $VirtG$ be the graph connecting all pairs of dominators $u$ and $v$ if there is a path in the original graph $G$ connecting them with at most 3 hops. It is well-known that the graph $VirtG$ is connected. It is natural to form a connected dominating set by finding connectors to connect any pair of dominators $u$ and $v$ if they are connected in $VirtG$. This strategy was used in several previous methods, such as [1, 3, 4, 5, 31].

Our new connector selection method for weighted connected dominating set is also based on this observation. First, we define two dominators $u$ and $v$ as *neighboring dominators* if they are at most 3 hops away, *i.e.*, they are neighbors in the graph $VirtG$. Let $\mathsf{LCP}(u, v, G)$ denote the least cost path $uv_1v_2 \cdots v_kv$ between vertices $u$ and $v$ on a weighted graph $G$, and $\mathcal{L}(u, v, G)$ denote the total cost of nodes on path $\mathsf{LCP}(u, v, G)$ excluding $u$ and $v$, *i.e.*, $\mathcal{L}(u, v, G) = \sum_{1 \le i \le k} c(v_k)$. For every pair of neighboring dominators $u$ and $v$, our method will find the shortest path with at most 3 hops to connect them. The nodes on this shortest path will be assigned a role of connector.

Our method uses the following data structures and messages.

1. $D_k(v)$ is the list of dominators that are $k$-hops away from a node $v$.
2. $P_k(v, u)$ is the least cost path from $v$ to $u$ using at most $k$-hops. Notice $u$ and $v$ may be less than $k$-hops away.
3. $\mathsf{OneHopDominatorList}(v, D_1(v))$: nodes $D_1(v)$ are the dominators of node $v$ that are 1-hop from $v$.
4. $\mathsf{TwoHopDominator}(v, u, w, c(w))$: node $u$ is a 2-hop dominator of node $v$ and the path $uwv$ has the least cost.

**Algorithm 2** Low-cost Connector Selection

1: Every dominatee node $v$ broadcasts to its 1-hop neighbors the list of its one-hop dominators $D_1(v)$ using message $\mathsf{OneHopDominatorList}(v, D_1(v))$. When a node $w$ receives $\mathsf{OneHopDominatorList}(v, D_1(v))$ from one-hop neighbor $v$, it puts the dominator $u \in D_1(v)$ to $D_2(w)$ if $u \notin D_1(w)$. Update the path $P_3(z, u)$ as $uvw$ if it has a smaller cost.
2: When a dominatee node $w$ received messages $\mathsf{OneHopDominatorList}$ from *all* its one-hop nodes, for each dominator node $u \in D_2(w)$, node $w$ sends out message $\mathsf{TwoHopDominator}(w, u, x, c(x))$, where $wxu$ is the least cost path $P_2(w, u)$.
3: When a dominator $z$ receives a message $\mathsf{TwoHopDominator}(w, u, x, c(x))$ from its neighbor $w$, it puts $u$ to $D_3(z)$ if $u \notin D_2(z)$, and updates the path $P_3(z, u)$ as $uwxz$ if $c(w) + c(x)$ has a less cost.
4: Each dominator $u$ builds a virtual edge $\widetilde{uv}$ to connect each neighboring dominator $v$. The length of $\widetilde{uv}$ is the cost of path $P_3(u, v)$. Notice that here the cost of end-nodes $u$ and $v$ is not included. All virtual edges forms an *edge weighted* virtual graph $VirtG$ in which all dominators are its vertices.
5: Run a distributed algorithm to build a MST on graph $VirtG$. Let $VMST$ denote $MST(VirtG)$.
6: For any virtual edge $e \in VMST$, select each of the dominatees on the path corresponding to $e$ as a connector.

The graph constructed by combining all of dominators and the connectors selected by the above algorithm is called a weighted connected dominating set (WCDS) graph (or *backbone*). Notice that since we run MST on graph $VirtG$, the constructed backbone is a sparse graph, i.e., it has only linear number of links.

# 4. PERFORMANCE GUARANTEE

In this section, we will first study the performances of the proposed weighted backbone structure in term of the total node cost in the backbone. Then, by a small modification of the backbone formation algorithm, we can make our weighted backbone more efficient for the unicast routing.

## 4.1 Total Cost of the Backbone

First, we would like to build a weighted backbone whose total node cost is as less as possible. We will show that the backbone constructed by our method is comparable to the optimum when the network is not dense, or the costs of the nodes do not have a dramatic change, *i.e.*, being smooth. Our analysis following is on the homogeneous networks, but it can be extended to general heterogeneous networks without difficulty. Before describing our result, we first review an important observation of the *dominating set* on UDG, which will play an important role in our proofs later. After clustering, one dominator node can be connected to many dominatees. However, it is well-known that a dominatee node can only be connected to at most *five* independent nodes in the unit disk graph model. In other words, the 1-*local independence number* of UDG, $\alpha^{[1]}(UDG)$, is 5. Generally, it is well-known that, for each node, there are at most a constant number ($\alpha^{[k]}(UDG)$) of independent nodes that are at most $k$ units away. The following lemma which bounds the number of independent nodes within $k$ units from a node $v$ is proved in [1] by using a simple area argument.

LEMMA 1. *For every node $v$, the number of independent nodes inside the disk centered at $v$ with radius $k$-units, $\alpha^{[k]}(UDG)$, is bounded by a constant $\ell_k = (2k + 1)^2$.*

The bounds on $\ell_k$ can be improved by a tighter analysis. In [40], Wan *et al.* gave the detailed proof to show that for unit disk graph the number of independent nodes in 2-hops neighborhood (not including the 1-hop neighbors) is at most 13 while the number of independent nodes in 1-hop neighborhood is at most 5. Therefore, there are at most 18 independent nodes inside the disk centered at a node $v$ with radius 2, *i.e.*, $\alpha^{[2]}(UDG) = 18$.

THEOREM 2. *Algorithm 1 constructs a dominating set whose total cost is no more than $\min(18 \log(\Delta + 1), 4\delta + 1)$ times of the optimum for networks modeled by UDG.*

PROOF. First, we prove the total cost of the maximal independent set $MIS$ formed by all $\mathsf{PossibleDominator}$ nodes is no more than $4\delta + 1$ times of the optimum. Assume node $u$ is a node from the optimum $OPT$. If $u$ is not a $\mathsf{PossibleDominator}$ node then there are at most 5 $\mathsf{PossibleDominator}$ nodes around $u$. Let $v_1^u, v_2^u, \cdots, v_5^u$ denote them. The cost of one of these five nodes is smaller than the cost of $u$, otherwise node $u$ will be selected as a $\mathsf{PossibleDominator}$ node. W.l.o.g., let $c(v_1^u) \le c(u)$. We also know that $c(v_i^u) \le \delta \cdot c(u)$ for $2 \le i \le 5$. Thus, $\sum_{1 \le i \le 5} c(v_i^u) \le (4\delta + 1)c(u)$. If we summarize the inequations for all nodes in the optimum dominating set $OPT$, we get

$$\sum_{u \in OPT} \sum_{1 \le i \le 5} c(v_i^u) \le (4\delta + 1) \sum_{u \in OPT} c(u) = (4\delta + 1)c(OPT).$$

Notice that every node in MIS will appear as $v_i^u$ for at least one node $u \in OPT$ since $OPT$ is a dominating set. Thus, $c(MIS) = \sum_{v \in MIS} c(v) \le \sum_{u \in OPT} \sum_{1 \le i \le 5} c(v_i^u)$. It follows that

$$c(MIS) \le (4\delta + 1)c(OPT).$$

Then, we prove the total cost of the nodes selected by the greedy method in Step 6 of Algorithm 1 is no more than $18 \log(\Delta + 1)$ times of the optimum. Assume that node $u$ runs the greedy algorithm and gets the subset as $GRDY_u$, and the cost of the selected

subset $c(GRDY_u)$ is at most $c(u)$. It is well known that the dominating set generated by the greedy algorithm for set cover is no more than $\log f$ times of the optimum if every set has at most $f$ items. Here, we know that every dominator can cover at most $\Delta$ dominatees, thus, $c(GRDY_u) \leq \log(\Delta+1) \cdot c(LOPT_u)$. Here $LOPT_u$ is an optimum dominating set (using nodes from $N_2(u)$) when the set of nodes to be covered are the 1-hop neighborhood of $u$ (including $u$). Assume that $OPT_u$ is the subset of the global optimum solution, denoted as OPT, for $MWCDS$ which falls in the 2-hops neighborhood of $u$, *i.e.*, $OPT_u = OPT \bigcap N_2(u)$. Obviously $OPT_u$ is a dominating set for $N_1(u)$. Thus, we have $c(LOPT_u) \leq c(OPT_u)$, since $LOPT_u$ is the local optimum. Therefore, $c(GRDY_u) \leq \log(\Delta+1) \cdot c(LOPT_u) \leq \log(\Delta+1) \cdot c(OPT_u)$. Consider all nodes in the MIS, we get

$$c(GRDY) \leq \sum_{u \in MIS} c(GRDY_u) \quad \leq \log(\Delta+1) \cdot \sum_{u \in MIS} c(OPT_u).$$

Remember that for each node $v$, the number of independent nodes in the 2-hops neighborhood of $v$ is bounded by 18. Therefore, each dominator is counted at most 18 times (once for each node $u \in MIS$ that selects $v$ to $GRDY_u$). Thus, $\sum_{u \in MIS} c(OPT_u) \leq 18c(OPT)$.

For each node $u$ in MIS, we either use $u$ as a dominator or use $GRDY_u$ as dominators, whichever has a smaller cost. Then, the total weight of the final dominating set is at most

$$\begin{aligned}
&\sum_{u \in MIS} \min(c(u), c(GRDY_u)) \\
&\leq \min(\sum_{u \in MIS} c(u), \sum_{u \in MIS} c(GRDY_u)) \\
&\leq \min(4\delta+1, 18\log(\Delta+1)) \cdot c(OPT).
\end{aligned}$$

This finishes our proof. □

Notice that here the approximation ratio is $\min(18\log(\Delta+1), 4\delta+1)$. So if one of $\log(\Delta+1)$ and $\delta$ is a constant, the approximation ratio is a constant. Our analysis is also pessimistic. As our simulation shows that the practical performance is much better than this theoretical bound. It is easy to generalize the above result to heterogeneous networks.

THEOREM 3. *For a network modeled by a graph $G$, Algorithm 1 constructs a dominating set whose total cost is no more than $\min(\alpha^{[2]}(G)\log(\Delta+1), (\alpha^{[1]}(G)-1)\delta+1)$ times of the optimum.*

Now, we need to prove the total cost of connectors selected by Algorithm 2 is also bounded. The following lemma about the relationship between $\mathcal{L}(u,v,G)$ and $\mathcal{L}(u,v,VirtG)$ will be used in the proof.

LEMMA 4. *For any pair of dominators $u$ and $v$,*

$$\mathcal{L}(u,v,VirtG) \leq 2 \cdot \mathcal{L}(u,v,G).$$

PROOF. Notice that the original graph is node weighted while the virtual graph $VirtG$ is edge weighted. Here, let $c(e)$ be the weight of edge $e = \widetilde{u_i u_j}$ and $c(e) = \mathcal{L}(u_i, u_j, G)$. We assume that path $uv_1v_2 \cdots v_k v$ is the least cost path connecting $u$ and $v$ in the original graph G, as shown in Figure 4.

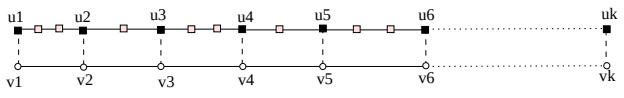



**Figure 4:** $\mathcal{L}(u,v,G) \geq 2 \cdot \mathcal{L}(u,v,VirtG)$**.**

For any dominatee node $p$ in original communication graph, it must be dominated by at least one dominator. Thus, we can assume that node $u_i$ is node $v_i$'s dominator as shown in Figure 4. For dominators $u_i$ and $u_{i+1}$, we argue that the length of $\widetilde{u_i u_{i+1}}$ is at most the summation of the cost of $v_i$ and $v_{i+1}$. Notice that $u_i v_i v_{i+1} u_{i+1}$ is a 3-hops path between $u_i$ and $u_{i+1}$ whose length is $c(v_i)+c(v_{i+1})$. Thus, the length of $\widetilde{u_i u_{i+1}}$ is at most $c(v_i)+c(v_{i+1})$. Thus we have $c(\widetilde{u_i u_{i+1}}) \leq c(v_i) + c(v_{i+1})$ for $1 \leq i \leq k-1$. Similarly, we also have $c(\widetilde{uu_1}) \leq c(v_1)$ and $c(\widetilde{u_k v}) \leq c(v_k)$. Summing all these inequalities, we get

$$\mathcal{L}(u,v,VirtG) \leq c(\widetilde{uu_1}) + c(\widetilde{v_k v}) + \sum_{i=1}^{k-1} c(\widetilde{u_i u_{i+1}}) \leq 2\sum_{i=1}^{k} c(v_i).$$

This finishes our proof. □

In graph $G$, we set all dominators' cost to 0 to obtain a new graph $G'$. Assume $T_{opt}$ is the tree with the minimum cost that spans all dominators selected by Algorithm 1. Following lemma shows that there exists a tree $T'_{opt}$ whose cost equals the cost of $T_{opt}$ and every dominatee node $u$ in $T'_{opt}$ has a node degree at most $\alpha^{[1]}(G)$.

LEMMA 5. *There exists a tree $T'_{opt}$ in $G'$ spanning all dominators selected in Algorithm 1 and connectors in this tree has degree at most $\alpha^{[1]}(G)$.*

PROOF. We prove this by construction. Consider any optimum cost tree $T_{opt}$ spanning all dominators. In tree $T_{opt}$, assume there exist some connectors whose degrees are greater than $\alpha^{[1]}(G)$. We choose any one of them as the root. The depth of a connector is defined as the hops from this connector to the root in $T_{opt}$. We process all connectors $u$ in $T_{opt}$ whose degree is greater than $\alpha^{[1]}(G)$ in an increasing order of their depths. Notice that, as we will see later, the depth of a node does change in our construction, but it will only increase. Assume that currently we are processing a node $u$ with more than $\alpha^{[1]}(G)$ neighbors. Clearly, there are at least two neighbors of $u$ in tree $T_{opt}$ that are connected, say $p$, $q$. Notice either $p$ or $q$'s depth is greater than $u$ since $u$ only has one parent. Without loss of generality, we assume that $p$'s depth is bigger than $u$'s depth. We then remove edge $uq$ and add edge $pq$. Then, $u$'s degree decreases by 1 while all other connectors whose depth is less than or equal to $u$'s remains unchanged and $p$'s degree increases by 1. Notice this will result in a new tree spanning all dominators while keep the cost of the tree unchanged. Update the depth of node $q$ and all nodes of the subtree rooted at $q$ (the depths will increase by one). Repeat the above iteration until all nodes are processed. It is obvious that the above process will terminate. The resulting tree is $T'_{opt}$. □

For tree $T'_{opt}$, we define its weight $c(T'_{opt})$ as the sum of the cost of all connectors. We also define $c(T) = \sum_{e \in T} c(e)$ for an edge weighted tree $T$. The above lemma implies that there is an optimum tree connecting all dominators with node degree at most 5 for networks modeled by UDG.

THEOREM 6. *The connectors selected by Algorithm 2 have a total cost no more than $2 \cdot \alpha^{[1]}(G)$ times of the optimum for networks modeled by $G$.*

PROOF. Let $K_G$ be another virtual complete graph whose vertices are all dominators selected in Algorithm 1 and edge length equal the cost of least cost path between two dominators on original graph $G$. Following we argue the weight of MST on graph $K_G$ is at most $\alpha^{[1]}(G)$ times the weight of tree $T'_{opt}$.

For spanning tree $T'_{opt}$, we root it at an arbitrary node and duplicate every link in $T'_{opt}$ (the resulting structure is called $DT'_{opt}$).

Clearly, every node in $DT'_{opt}$ has an even degree now. Thus, we can find an Euler circuit, denoted by $EC(DT'_{opt})$, that uses every edge of $DT'_{opt}$ exactly once, which is equivalent to say that every edge in $T'_{opt}(G)$ is used exactly twice. Consequently, every node $v_k$ in $V(T'_{opt})$ is used exactly $d_{T'_{opt}}(v_k)$ times. Here $d_G(v)$ denotes the degree of a node $v$ in a graph $G$. Thus, the total weight of the Euler circuit is at most $\alpha^{[1]}(G)$ times of $c(T'_{opt})$, *i.e.*,

$$c(EC(DT'_{opt})) \leq \alpha^{[1]}(G) \cdot c(T'_{opt}).$$

Notice that here if a node $v_k$ appears multiple times in $EC(DT'_{opt})$, its weight is also counted multiple times in $c(EC(DT'_{opt}))$.

If we walk along $EC(DT'_{opt})$, we visit all dominators, and the length of any subpath between dominators $u_i$ and $u_j$ is not smaller than $\mathcal{L}(u_i, u_j, G)$. Therefore, the cost of $EC(DT'_{opt})$ is at least $c(MST(K_G))$ since $MST(K_G)$ is the minimum cost tree spanning all dominators and the edge $u_iu_j$ in $MST(K_G)$ corresponds to the path with the least cost between $u_i$ and $u_j$. In other words,

$$c(EC(DT'_{opt})) \geq c(MST(K_{UDG})).$$

Consequently, we have

$$c(MST(K_G)) \leq c(EC(DT'_{opt})) \leq \alpha^{[1]}(G) \cdot c(T'_{opt}). \quad (1)$$

Now we prove the weight of $MST(VirtG)$ is at most two times the weight of $MST(K_G)$. For any edge $e = u_iu_j \in MST(K_G)$, from Lemma 4, we have

$$c(e) \geq \mathcal{L}(u_i, u_j, G) \geq \frac{\mathcal{L}(u_i, u_j, VirtG)}{2}.$$

For each edge $e = u_iu_j \in MST(K_G)$, we connect them in graph $VirtG$ using path $\mathsf{LCP}(u_i, u_j, VirtG)$. This constructs a connected subgraph $MST'$ on graph $VirtG$ whose cost is not greater than twice of the weight of $MST(K_G)$. Thus, we have

$$c(MST(VirtG)) \leq c(MST') \leq 2 \cdot c(MST(K_G)). \quad (2)$$

The theorem follows from combining inequalities (1) and (2): $c(MST(VirtG)) \leq 2c(MST(K_G)) \leq 2\alpha^{[1]}(G) \cdot c(T'_{opt})$. □

Notice that Theorem 6 also implies the following side-product result: given a group of receivers in a node weighted network, the connectors found through VMST have total cost no more than $2\alpha^{[1]}(G)$ times of the minimum cost multicast tree. For the special case of UDG, the total cost of the connectors is no more than 10 times of the optimum multicast tree. Here we assume that the receivers have cost 0.

Combining Theorem 3 and Theorem 6, we get the following theorem which is one of the main contributions of this paper.

THEOREM 7. *For any communication graph $G$, our algorithm constructs a weighted connected dominating set whose total cost is no more than*

$$\min(\alpha^{[2]}(G)\log(\Delta+1), (\alpha^{[1]}(G)-1)\delta+1) + 2\alpha^{[1]}(G)$$

*times of the optimum.*

Specifically, when the networks are modeled by a unit disk graph, we have the following corollary.

COROLLARY 8. *For homogeneous wireless networks, our algorithm constructs a weighted connected dominating set whose total cost is no more than* $\min(18\log(\Delta+1), 4\delta+1)+10$ *times of the optimum.*

## 4.2 Unicast Performance

After we construct the backbone WCDS, if a node $u$ wants to broadcast a message, it follows the following procedure. If node $u$ is not a dominator, then it sends the message to one of its dominators. When the message reaches the backbone, it will be broadcast along the virtual minimal spanning tree. In previous section, we prove that the total cost of WCDS is no more than a constant times of the optimum, which implies that our structure is energy efficient for broadcast.

When considering unicast routing, we can modify our backbone formation algorithms by

1. removing steps 5, 6, and 7 (collecting 2-hop information and running the greedy algorithm for set over) from Algorithm 1;
2. modifying PossibleDominator to Dominator in step 2 of Algorithm 1; and
3. removing steps 5 and 6 (building VMST) from Algorithm 2.

Notice that the changes to Algorithm 1 are not necessary as will see later. Let $UWCDS$ be the constructed backbone. If a node $u$ wants to unicast a message, it follows the following procedure. If node $u$ is not a dominator and node $v$ is not a neighbor of $u$, node $u$ sends the message to one of its dominators. Then the dominator will transfer the message to the target or a dominator of the target through the backbone. Now, we prove that the backbone is a spanner for unicast application, *i.e.*, every route in the constructed network topology is efficient. Remember a route is *efficient* if its total cost (or total hop number) is no more than a constant factor of the minimum total cost (or total hop number) needed to connect the source and the destination in the original communication graph. The constant is called cost (or hops) stretch factor.

We first prove the backbone has a bounded cost stretch factor.

THEOREM 9. *For any communication graph, the cost stretch factor of UWCDS is at most* 3.

PROOF. Consider any source node $s$ and target node $t$ that are not connected directly in the original communication graph $G$. Assume the least cost path $\mathsf{LCP}(s, t, G)$ from $s$ to $t$ in $G$ is $\Pi_{G_h}(s, t) = v_1v_2...v_k$, where $v_1 = s$ and $v_k = t$, as illustrated by Figure 4. We construct another path in UWCDS from $s$ to $t$ and the total cost of this path is at most 3 times of the cost of the least cost path $\mathsf{LCP}(s, t, G)$.

For any dominatee node $p$ in original communication graph $G$, we will show that there must exist one dominator $q$ whose cost is not greater than $p$'s cost. First, from our selection procedure of the maximal independent set, node $p$ is not selected to MIS implies that, at some stage, there is a neighbor, say $u$, with smaller cost selected to MIS, which will be PossibleDominator. Notice that, this PossibleDominator node $u$ may not appear in our final structure. However, this node is not selected only if $c(GRDY_u)$ is smaller than $c(u)$. Notice that clearly, there is at least one node, say $v$, in $GRDY_u$ that dominates node $p$ since $p$ is a one-hop neighbor of node $u$ and $GRDY_u$ covers all one hop neighbors of $u$ (including $u$). Clearly, all dominators in $GRDY_u$ has cost no more than $c(u)$ from $c(GRDY_u) \leq c(u)$. If node $u$ is in final structure, we set $q$ as $u$, otherwise, set $q$ as node $v$. We call node $q$ as node $p$'s *small dominator*. Notice that $q$ and $p$ can be the same node.

For each node $v_i$ in the path $\mathsf{LCP}(s, t, G)$, let $u_i$ be its small dominator if $v_i$ is not a dominator, else let $u_i$ be $v_i$ itself. Notice that there is a 3-hop path $u_iv_iv_{i+1}u_{i+1}$ in the original communication graph $G$. Then from Algorithm 2, we know there must exist one or two connectors connecting $u_i$ and $u_{i+1}$, and also the cost summation of these connectors is at most the cost summation of $v_i$ and $v_{i+1}$. We define a path, denoted by $\mathsf{LCP}(s, t, UWCDS)$, to connect $s$ and $t$ in UWCDS as the concatenation of all paths

$\mathsf{LCP}(u_i, u_{i+1}, VirtG)$, for $1 \leq i \leq k-2$, and a least cost path (with $\leq$ two hops) connecting $u_{k-1}$ and $t$. Remember that the path $\mathsf{LCP}(u_i, u_{i+1}, VirtG)$ is only the least cost path among all paths connecting $u_i$ and $u_{i+1}$ using at most 3 hops.

We then show that the path $\mathsf{LCP}(s, t, UWCDS)$ has a cost no more than 3 times of the path $\mathsf{LCP}(s, t, G)$, where $\mathsf{LCP}(s, t, G)$ is the least cost path connecting $s$ and $t$ in the original communication graph $G$. Clearly, $\sum_{i=1}^{k-2} \mathcal{L}(u_i, u_{i+1}, VirtG) \leq c(v_1) + 2 \cdot \sum_{i=2}^{k-2} c(v_i) + c(v_{k-1})$. Notice that, in our unicast routing algorithm, when the target node $t$ is within two hops of the dominator node $u_{k-1}$, node $u_{k-1}$ will not send the data to dominator node $u_k$. Instead, if target $t$ is one hop neighbor of node $u_{k-1}$, it will directly send data to node $t$; otherwise, node $u_{k-1}$ will find a least cost node, say $w$, to connect to the target node $t$ directly. Obviously, $c(w) \leq c(v_{k-1})$ since node $v_{k-1}$ connects $u_{k-1}$ and target $t$. Thus, the total cost of the path in the constructed backbone is

$$\sum_{i=1}^{k-2} \mathcal{L}(u_i, u_{i+1}, VirtG) + \mathcal{L}(u_{k-1}, t, VirtG) + \sum_{i=1}^{k-1} c(u_i)$$

$$\leq c(v_1) + 2\sum_{i=2}^{k-2} c(v_i) + c(v_{k-1}) + c(v_{k-1}) + \sum_{i=1}^{k-1} c(v_i) < 3\sum_{i=1}^{k-1} c(v_i).$$

This finishes our proof. □

Similar with the proof in [1], we can prove that

THEOREM 10. *For any communication graph (not necessarily a UDG model), the hops stretch factor of UWCDS is at most* 4.[2]

## 4.3 Message Complexity

Compared with data processing, wireless node spends more energy in data communication. Here we show that our algorithms are efficient in term of communication complexity.

THEOREM 11. *Algorithm 1 uses $O(n)$ messages if the networks are modeled by UDG and the geometry information of all nodes is known.*

PROOF. First, for messages ItryDominator and IamDominatee, every node at most sends out once this kind of messages. Thus, the total number of these two messages is $O(n)$.

Second, for each PossibleDominator node, it needs to collect the costs and IDs of all of its two hop neighbors. This step may cost lots of communications (at most $O(m)$ messages when no geometry information is known, where $m$ in the number of links in the original UDG). Recently Calinescu [11] proposed a communication efficient method (using $O(n)$ messages) to collect $N_2(u)$ for every node $u$ when the geometry information is known for networks modeled by UDG.

Third, after applying the greedy method node $u$ may send a message YouAreDominator to node $v$, but since the number of independent nodes $u$ in two hops of $v$ is bounded by a constant, the total number of this kind of messages is also $O(n)$.

Consequently, Algorithm 1 uses $O(n)$ messages. □

It is easy to show that Algorithm 1 uses $O(m)$ messages for a general networks or the geometry information of all nodes is unknown. For Algorithm 2, first, the number of messages in the first three steps is at most $O(m)$. Obviously, we can construct the minimum spanning tree on $VirtG$ using $O(m + n \log n)$ number of messages. In practice, we may not need construct the minimum spanning tree exactly: a localized approximation of the minimum spanning tree [29] may perform well enough, which has a message complexity only $O(n)$. In addition, if only unicast running on the backbone, we can ignore the MST construction, then the message complexity is only $O(m)$.

[2]Actually, the bound is $3 + \frac{2}{k}$, where $k$ is the number of hops of the shortest hop path in the original communication graph. The basic idea of the proof is similar with the idea used in proof of Lemma 4 and illustrated by the example in Figure 4. Since 1-hop neighbors can directly communicate with each other, for any nodes that are at least 2-hops away, the bound is 4.

## 4.4 Time Complexity

Considering the data processing at each wireless node, we also study the time complexity of our algorithms.

For Algorithm 1, the first four steps take at most $O(n)$ in time. To collect the information of two-hop neighbors, we apply the method proposed by Calinescu [11], which also takes at most $O(n)$ in time. Notice that the time complexity of the greedy method in [15, 16] (based on the set covering method in [14]) is at most $O(m\Delta)$, where $m$ is number of nodes participating in the algorithm and $\Delta$ is the maximum node degree. So the sixth step of Algorithm 1 takes at most $O(\Delta_2\Delta)$ where $\Delta_2$ is the maximum number of two-hop neighbors. Since $\Delta_2 \leq n$ and $\Delta_2 \leq \Delta^2$, the sixth step takes at most $O(\Delta^3)$ (or $O(n\Delta)$). Therefore, the time complexity of Algorithm 1 is $O(n\Delta)$ in worst case.

For Algorithm 2, the most time consuming step is to build a MST on $VirtG$. Obviously, we can construct the MST using at most $O(m + n \log n)$ time.

# 5. DISCUSSIONS

## 5.1 Practical Applications

As we mentioned in the introduction (Section 1), the proposed distributed algorithms can be used in wireless ad hoc networks to form a low-cost network backbone for unicast routing or broadcasting application. The cost which we used as the input of our algorithms could be a *generic* cost, which defined by various practical applications. Here we list some possible weights maybe used in wireless ad hoc networks.

**Energy Consumption Rate**: Most backbone-based unicast routing or broadcasting protocols [15, 30, 35, 41] deliver packets only through the backbone or restrict the flooding packets in the backbone, thus the nodes serving as clusterheads or connectors in the backbone consume more energy than ordinary nodes. If we use the energy consumption rate at each node as its weight, using the proposed low-cost backbone formation algorithm, we can achieve an energy efficient backbone where the total energy consumption of this backbone is at most constant times of the energy consumption of the optimum. Also the unicast carried on the backbone is also power efficient, compared with the least energy consumption path in the original communication graph.

Another way to build energy-efficient backbone is to select nodes with the maximum amount of remaining energy (equivalently, the minimum amount of consumed energy if the initial energy of each node is same).

**Fault Tolerant Rate**: Fault tolerance is also an important issue in wireless ad hoc networks, since nodes are mobile and in a dynamic environment. If each node estimates its probability of being fault and we treat it as the weight, we can use our algorithm to build a fault-tolerant backbone for routing. The fault tolerant rate can be evaluated by considering the mobility (stability, speed) of the node, the quality of links (link failures) around the node, the interference level at the node, or other metric. Some research along this line have been done in [7, 10, 24, 34]. Assume that $p_i$ is the probability that the wireless node $v_i \in V$ will have fault in computing or communicating with its neighbors. Two possible criteria could be used to measure the fault-tolerant quality of a backbone (*i.e.*, a

CDS $S \subset V$): $\sum_{v_i \in S} p_i$ or $\Pi_{v_i \in S} p_i$. In the first case, the cost (or called weight) of node $v_i$ is assigned as $c(v_i) = p_i$, while in the latter case, the cost of $v_i$ is assigned as $c(v_i) = \log p_i$. Then building most fault-tolerant backbone is equivalent to find a CDS with the minimum total cost.

**Security Level**: Our algorithm can also be applied in designing secure routing protocols. Since ad hoc networks lack a central authority for authentication and key distribution, security is hard to achieve. In [33], Liu *at el.* proposed a dynamic trust model for ad hoc network. Each node has a security level by observing its neighbor. By using the security level information got from their method, we can apply our low-cost method to build a backbone for routing with high security. We could assign the cost to a node using a method analog to the case of fault-tolerance discussed above.

More different metrics can be considered as the weight in our method, such as traffic load, signal overhead, battery level, and coverage. As done in [12, 13, 38], we can also use a combined weight function to integrate various metrics in consideration to form a more robust and efficient backbone for wireless ad hoc networks in general applications.

Beside forming the backbone for routing or broadcasting, our cluster algorithm (Algorithm 1) can also be used in other applications. For example, Zheng *et al.* [42] studied the *time indexing* problem in sensor networks. To enable time-indexed in-network storage of sensor data, they selected a subset of sensors, *i.e.*, rendezvous points to collect, compress and store sensor data from its neighborhood for pre-defined periods of time. To consider the energy and storage balancing, we can apply our weighted cluster algorithm to select the rendezvous points. Another example, in [23], a simple cluster algorithm is used for selecting the mobile agents to perform intrusion detection in wireless ad hoc networks. We can also apply our method to their intrusion detection system to achieve more robust and power efficient agent selection.

## 5.2 Dynamic Update

After the generation of the weighted backbone, dynamic maintenance of the backbone is also an important issue, since an ad hoc network could be highly dynamic. Two major events may cause the backbone obsoleted: 1) *topology changes* due to node moving, node joining or leaving, node failure; and 2) *weight changes* when weights are assigned based on some observed status of nodes. Notice that some of the practical weights we discussed above change frequently, such as battery level and quality of links. Therefore, a dynamic update method for our backbone is needed. Usually, there are two kinds of update methods: on-demand update or periodical update. Most of the existing clustering algorithms are invoked periodically, while some algorithms (such as [12]) perform the updating only when it is required (*i.e.*, on-demand). Our algorithm can adapt and combine both of these two update methods. If no major topology change or no remarkable weight change, no update will be performed until some pre-set timer expires. In other words, we perform our algorithm periodically with a pre-set time. The time could be set quite long depending on the types of the weight and applications. This kind of global update also insures the load balance throughout the network. But for some major topology change (such as a clusterhead dies) or tremendous change of weights (such as a big drop of security level), an on-demand update will be performed. Notice that since our algorithm is a localized algorithm[3], the update process can be performed only in a local area where the change occurs. However, it remains an open problem how to update the topology efficiently while preserving the approximation quality.

[3]By using *localized minimum spanning tree* (LMST) instead of MST, our distributed algorithm becomes a localized algorithm. We will discus it in Section 6.1 in detail.

# 6. PERFORMANCE EVALUATION

In this section, we conduct extensive simulations to study the performances of our proposed backbone and compared them with previously greedy algorithms.

## 6.1 Practical Implementation

Since the distributed construction of MST in Algorithm 2 is expensive, we implement a localized approximation of MST, *localized minimum spanning tree* (LMST) [29]. For completeness, we define LMST for general edge weighted graph $G$ here.

DEFINITION 1. *The $k$-local minimum spanning tree ($LMST_k$) contains a* directed *edge $\overrightarrow{uv}$ if edge $uv$ belongs to $MST(N_k(u))$.*

For the edge weighted graph $VirtG$, each dominator node $u$ will first collect all dominator nodes that are at most $k$-hops away in $VirtG$. Typically $k$ is 1 or 2 in our methods. Node $u$ then constructs the minimum spanning tree $MST(N_k(u))$ and keep all edges $uv \in MST(N_k(u))$. The union of all such selected links are called the local minimum spanning tree, denoted by $LMST_k(G)$. Notice that here the weight of a link $uv$ is the cost of the least cost path (with $\leq 3$ hops) connecting $u$ and $v$ in $G$. From the property of the minimum spanning tree, the following lemma is obvious.

LEMMA 12. *The global minimum spanning tree $MST(G)$ is a subgraph of the local minimum spanning tree $LMST_k(G)$.*

Unfortunately, in the worst case, the total cost of $LMST_k(G)$ could be arbitrarily larger than the cost of $MST(G)$. However, our simulations show that it is within a small constant factor on average. The advantage of using the local minimum spanning tree instead of the global minimum spanning tree is the significant reduction in the communication cost.

## 6.2 Performance Comparisons

We then evaluate the performance of our new distributed weighted backbone formation algorithm by simulations on random networks. In our experiments, we randomly generated a set $V$ of $n$ wireless nodes with random costs drawn from $[1, 100]$ and the induced $UDG(V)$, then tested the connectivity of $UDG(V)$. If it is connected, we construct different cluster algorithms on $UDG(V)$ to form dominating sets and measure the total costs of these dominating sets. Then, we apply our new method to construct the weighted backbone. We test the total cost of the final backbone and measure the average and maximum cost/hop spanning ratios.

In the experimental results presented here, $n$ wireless nodes are randomly distributed in a $500m \times 500m$ square, and the transmission range is set to $100m$. We tested all algorithms by varying $n$ from 50 to 275, where 50 vertex sets are generated for each case to smooth the possible peak effects. The average and the maximum were computed over all these 50 vertex sets. Notice, the parameter setting of our experiments here is just for demonstrations. We have tried other various settings, the results and performances are stable, due to space limit, we can not present all of them here.

### *6.2.1 Cost of Dominators*

First, we compare our algorithm with the three previous greedy algorithms to find a dominating set. Figure 5 gives an example of the original communication graph with node costs and different dominating sets by different greedy methods.

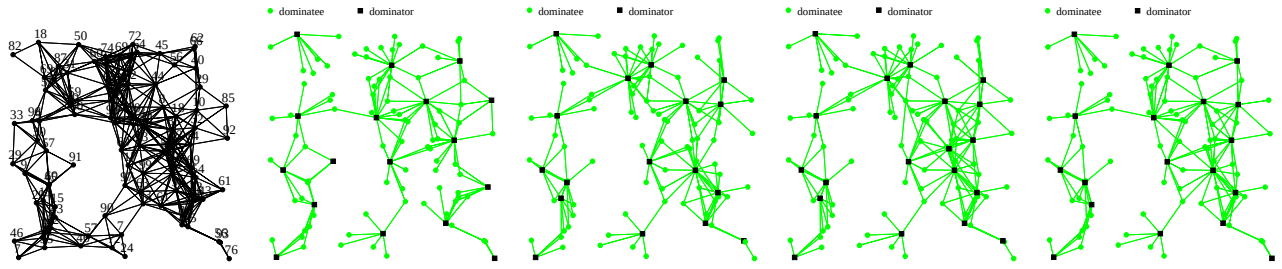



**Figure 5: Different dominating sets by different greedy methods from the same original communication graph.**

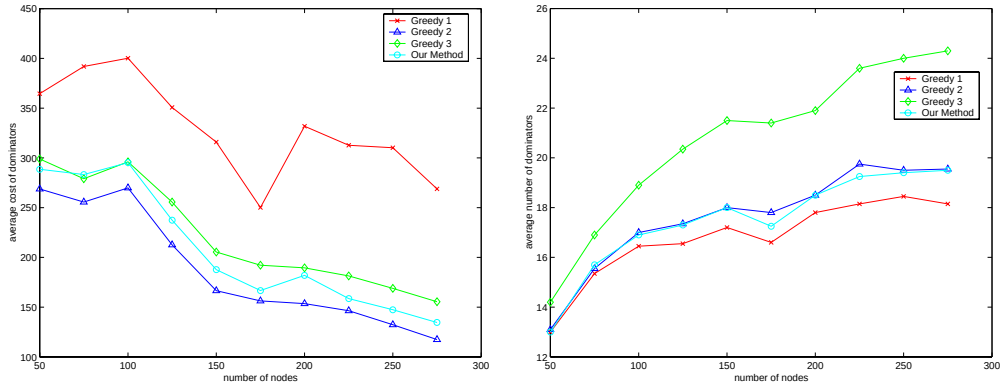



**Figure 6: Total cost and number of cluster-head of different greedy methods (when the number of nodes are from $50$ to $275$).**

We plotted the performances of all methods in Figure 6. Our method produces a dominating set whose cost is significantly less than that produced by the MIS based method (greedy 1) and is on the similar level with other two methods. In addition, our method produces a dominating set whose size is significantly less than that produced by the method in [6] (greedy 3) and is on the similar level with other two methods. The set-cover based method (greedy 2) is the only one that is comparable with our method for both metrics. However, it is a centralized method while our is a distributed method with a small communication cost.

### 6.2.2 *Cost of Backbone*

After getting the dominating set (Figure 7(b)) by Algorithm 1, we apply Algorithm 2 to find the connectors. Figure 7(c) shows the backbone after adding some connectors to the dominating set. Notice that we used the local minimum spanning tree to find the connectors instead of the global minimum spanning tree (that is why the graph WCDS in Figure 7(c) is not a tree). We plot the total cost and the size of the weighted backbone in Figure 8 (a) and (b). The size of the backbone becomes stable when the network becomes denser. However, the average total cost of the backbone decreases over the increasing of the network density, which is due to dense network provides more candidates for backbone with potential lower costs.

### 6.2.3 *Cost of Unicast Routing*

For unicast, we can simplify Algorithm 2 by directly using VirtG as the final backbone. Figure 7(d) illustrates such backbone. Spanning ratios of the final backbone are plotted in Figure 8 (c). Notice that the average cost and hop spanning ratios are indeed small (almost 1). The maximum cost spanning ratio is less than 3. The maximum hop spanning ratio is no more than 4. These maps well to the theoretical bounds, which are 3 and 4 respectively.

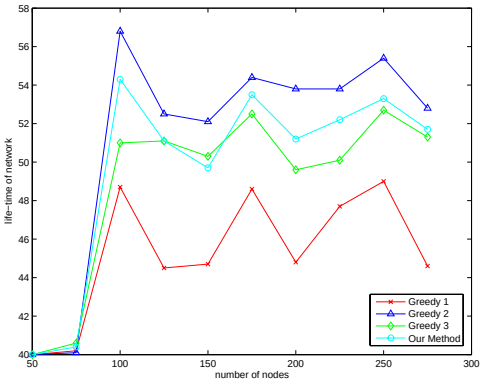
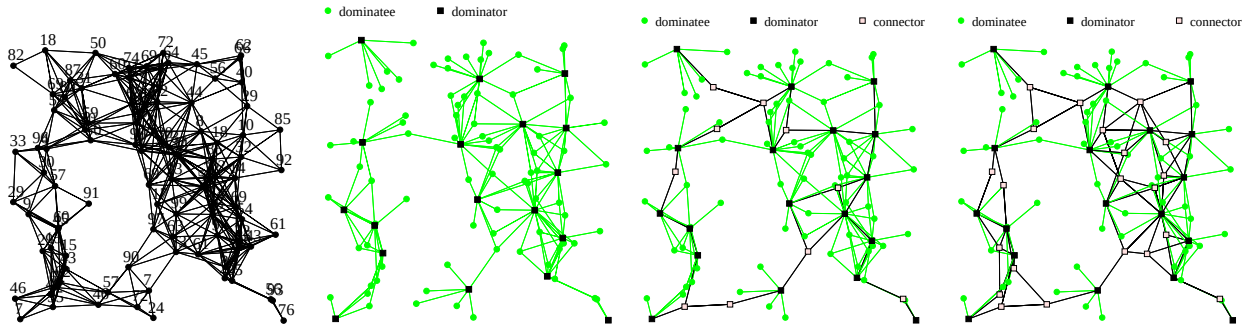



**Figure 9: The life time of the network using different greedy methods (when the number of nodes are from $50$ to $275$).**

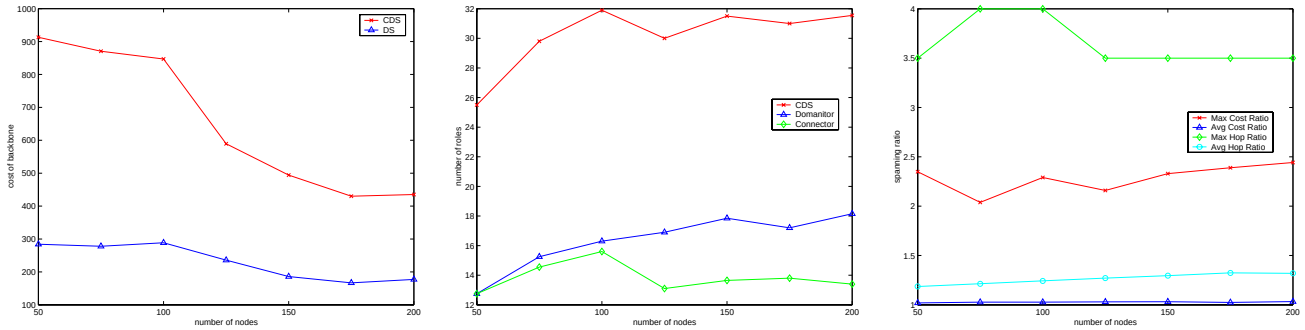
### 6.2.4 *Life-time Experiments*

We also conduct simple experiments to test the life time of the network when using our proposed backbone. Using the same random distribution and transmission range as in previous experiments, we setup networks in a $500m \times 500m$ square. Then, we assume that each node in the network has total energy $200$ initially. We perform the three classical greedy cluster algorithms and our algorithm to build weighted backbone for the networks and update it periodically. To ignore the effects of methods selecting connectors, we apply the same method used in our solutions (Algorithm 2) to




(a) UDG (b) DS (c) WCDS (d) VirtG

**Figure 7: Dominating set, connected dominating set and virtual backbone for unicast from the same original communication graph.**




(a) total cost (b) total number (c) spanning ratio

**Figure 8: Performance of backbone (when the number of nodes are from $50$ to $200$).**

connect clusters generated by the different cluster algorithms. For the cost of each node being backbone, we simply use the reverse of the remaining energy at each node. At the end of each period, we reduce the power of backbone nodes by $5$, and update the backbone. Figure 9 shows the life time (the number of periods that the network survives until the first node run out of energy). Again, our method has longer life time than the MIS based method (greedy 1) and is on the similar level with other two methods. Remember the set-cover based method (greedy 2) is a centralized method, therefore it has good performance in this experiment. Notice, the third greedy algorithm also has similar performance with our method. The reason maybe as follows. Even the size of the dominating set generated by greedy 3 is larger than our method (as shown in the first experiment), after selecting the connectors the size of backbone is in same level with our method for a random distributed network.

## 7. SUMMARY AND FUTURE WORK

In this paper, we present a new algorithm to construct a sparse structure for network backbone in wireless ad hoc networks. A communication efficient distributed algorithm was presented for the construction of a weighted connected dominating set, whose size is guaranteed to be within a small constant factor of the minimum (when either $\delta$ or $\Delta$ is a constant). We also show that with a small modification the constructed backbone is efficient for both cost and hops (though losing the low cost property). This topology can be constructed locally and is easy to maintain when the nodes move around. Our simulations confirmed that our new backbone indeed has well performances in random networks.

There are many interesting open problems left for further study. Remember that, we use the following assumptions on wireless network model: omni-directional antenna, single transmission received by all nodes within the vicinity of the transmitter, nodes being static for a reasonable period of time. To prove that the backbone has low cost, we also assume that all nodes have the same transmission range. Notice that the efficiency property for unicast does not require the communication graph to be a UDG. The problem will become much more complicated if we relax some of these assumptions. Another interesting open problem is to study the dynamic updating of the backbone efficiently when nodes are moving in a reasonable speed although our cost function may integrate the mobility of the nodes. It is interesting to see the practical performance differences of all proposed methods such as methods by Baker *et al.*, Alzoubi *et al.*, and our methods proposed here, in mobile environment.

## 8. REFERENCES

[1] ALZOUBI, K., LI, X.-Y., WANG, Y., WAN, P.-J., AND FRIEDER, O. Geometric spanners for wireless ad hoc networks. *IEEE Trans. on Paral. and Distr. Processing 14*, 4 (2003), 408–421.

[2] ALZOUBI, K. M., WAN, P.-J., AND FRIEDER, O. Message-optimal connected dominating sets in mobile ad hoc networks. In *Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking & computing* (2002), ACM Press, pp. 157–164.

[3] ALZOUBI, K. M., WAN, P.-J., AND FRIEDER, O. New distributed algorithm for connected dominating set in wireless ad hoc networks. In *HICSS, Hawaii* (2002).

[4] BAKER, D., EPHREMIDES, A., AND FLYNN, J. A. The Design and Simulation of a Mobile Radio Network with Distributed Control. *IEEE Journal on Selected Areas in Comm.* (1984), 226–237.

[5] BAKER, D. J., AND EPHREMIDES, A. The Architectural Organization of a Mobile Radio Network via a Distributed Algorithm. *IEEE Transactions on Communications COM-29*, 11 (November 1981), 1694–1701.

[6] BAO, L., AND GARCIA-LUNA-ACEVES, J. J. Topology management in ad hoc networks. In *Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing* (2003), ACM Press, pp. 129–140.

[7] BASAGNI, S. Distributed clustering for ad hoc networks. In *Proceedings of the IEEE Intern. Symp. on Parallel Architectures, Algorithms, and Networks* (1999), pp. 310–315.

[8] BASAGNI, S. Finding a maximal weighted independent set in wireless networks. *Telecommunication Systems, Special Issue on Mobile Computing and Wireless Networks 18*, 1/3 (September 2001), 155–168.

[9] BASAGNI, S., MASTROGIOVANNI, M., AND PETRIOLI, C. A performance comparison of protocols for clustering and backbone formation in large scale ad hoc networks. In *1st IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS)* (2004).

[10] BETTSTETTER, C., AND KRAUSSER, R. Scenario-based stability anlysis of the distributed mobility-adaptive clustering (dmac) algorithm. In *Proceedings of the 2nd ACM international symposium on Mobile ad hoc networking & computing* (2001), ACM Press, pp. 232–241.

[11] CĂLINESCU, G. Computing 2-hop neighborhoods in ad hoc wireless networks. In *AdHoc-Now 03* (2003).

[12] CHATTERJEE, M., DAS, S., AND TURGUT, D. WCA: A weighted clustering algorithm for mobile ad hoc networks. *Journal of Cluster Computing 5*, 2 (2002), 193–204.

[13] CHEN, G., NOCETTI, F., GONZALEZ, J., AND STOJMENOVIC, I. Connectivity-based k-Hop clustering in wireless networks. In *Proceedings of the 35th HICSS-Volume 7* (2002), IEEE Computer Society, p. 188.3.

[14] CHVÁTAL, V. A greedy heuristic for the set-covering problem. *Mathematics of Operations Research 4*, 3 (1979), 233–235.

[15] DAS, B., AND BHARGHAVAN, V. Routing in ad-hoc networks using minimum connected dominating sets. In *1997 IEEE Intern. Conference on on Communications (ICC'97)* (1997), vol. 1, pp. 376–380.

[16] DAS, B., SIVAKUMAR, R., AND BHARGHAVAN, V. Routing in ad hoc networks using a spine. In *IEEE Sixth International Conference on Computer Communications and Networks (ICCCN97)* (1997).

[17] GAO, J., GUIBAS, L. J., HERSHBURGER, J., ZHANG, L., AND ZHU, A. Discrete mobile centers. In *Proceedings of the 17th ACM Symposium on Computational Geometry (SoCG 01)* (2001), pp. 188–196.

[18] GERLA, M., AND TSAI, J. T.-C. Multicluster, mobile, multimedia radio network. *Wireless Networks 1*, 3 (1995), 255–265.

[19] GUHA, S., AND KHULLER, S. Approximation algorithms for connected dominating sets. *Algorithmica 20*, 4 (1998), 374–387.

[20] GUHA, S., AND KHULLER, S. Improved methods for approximating node weighted steiner trees and connected dominating sets. *Information and Computation 150*, 1 (1999), 57–74.

[21] HEINZELMAN, W. R., CHANDRAKASAN, A., AND BALAKRISHNAN, H. Energy-efficient communication protocol for wireless microsensor networks. In *Proceedings of the 33rd HICSS-Volume 8* (2000), IEEE Computer Society, p. 8020.

[22] HUNT, H. B., MARATHE, M. V., RADHAKRISHNAN, V., RAVI, S. S., ROSENKRANTZ, D. J., AND STEARNS, R. E. NC-Approximation Schemes for NP- and PSPACE -Hard Problems for Geometric Graphs. *Journal of Algorithms 26*, 2, (1998), 238-274.

[23] KACHIRSKI, O., AND GUHA, R. Intrusion detection using mobile agents in wireless ad hoc networks. In *Proceedings of IEEE Workshop on Knowledge Media Networking/* (2002).

[24] KOZAT, U. C., KONDYLIS, G., RYU, B., AND MARINA, M. Virtual dynamic backbone for mobile ad hoc networks. In *Proceedings of IEEE ICC 2001* (2001).

[25] KUHN, F., MOSCIBRODA, T., AND WATTENHOFER, R. What cannot be computed locally. In *Proceedings of the 23rd annual symposium on Principles of distributed computing* (2004), ACM Press, pp. 300–309.

[26] KUHN, F., MOSCIBRODA, T., AND WATTENHOFER, R. Initializing newly deployed ad hoc and sensor networks. In *Proceedings of the 10th annual international conference on Mobile computing and networking (MobiCom)* (2004), ACM Press, pp. 260–274.

[27] KUHN, F., AND WATTENHOFER, R. Constant-time distributed dominating set approximation. In *Proceedings of the 22nd annual symposium on Principles of distributed computing* (2003), ACM Press, pp. 25–32.

[28] LI, X.-Y., AND WANG, Y. Simple heuristics and PTASs for intersection graphs in wireless ad hoc networks. In *ACM 6th International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications (DIALM2002)* (2002).

[29] LI, X.-Y., WANG, Y., WAN, P.-J., SONG, W.-Z., AND FRIEDER, O. Localized low weight graph and its applications in wireless ad hoc networks. In *IEEE INFOCOM 2004* (2004).

[30] LIANG, B., AND HAAS, Z. J. Virtual backbone generation and maintenance in ad hoc network mobility management. In *Proc. of 19th IEEE INFOCOM* (2000), vol. 3, pp. 1293–1302.

[31] LIN, C. R., AND GERLA, M. Adaptive clustering for mobile wireless networks. *IEEE Journal of Selected Areas in Communications 15*, 7 (1997), 1265–1275.

[32] LIU, H., AND GUPTA, R. Selective backbone construction for topology control. In *1st IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS)* (2004).

[33] LIU, Z., JOY, T., AND THOMPSON, R. A dynamic trust model for mobile ad hoc networks. In *Proceedings of the 10th IEEE International Workshop on Future Trends in Distributed Computing Systems* (2004).

[34] MIN, M., WANG, F., DU, D.-Z., AND PARDALOS, P. M. A reliable virtual backbone scheme in mobile ad-hoc networks. In *1st IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS)* (2004).

[35] SIVAKUMAR, R., DAS, B., AND BHARGHAVAN, V. The clade vertebrata: spines and routing in ad hoc networks. In *IEEE Symposium on Computers and Communications* (1998).

[36] SMARAGDAKIS, G., MATTA, I., AND BESTAVROS, A. SEP: A stable election protocol for clustered heterogeneous wireless sensor networks. In *Second International Workshop on Sensor and Actor Network Protocols and Applications* (2004).

[37] STOJMENOVIC, I., SEDDIGH, M., AND ZUNIC, J. Dominating sets and neighbor elimination based broadcasting algorithms in wireless networks. *IEEE Transactions on Parallel and Distributed Systems 13*, 1 (2002), 14–25.

[38] TURGUT, D., DAS, S., ELMASRI, R., AND TURGUT, B. Optimizing clustering algorithm in mobile ad hoc networks using genetic algorithmic approach. In *IEEE GLOBECOM 2002* (2002).

[39] WAN, P.-J., ALZOUBI, K. M., AND FRIEDER, O. Distributed construction of connected dominating set in wireless ad hoc networks. In *INFOCOM* (2002).

[40] WAN, P.-J., LI, X.-Y., AND SONG, W.-Z. Theoretically good distributed OVSF-CDMA code assignment in wireless ad hoc networks, 2004. Submitted for publication.

[41] WU, J., AND LI, H. A dominating-set-based routing scheme in ad hoc wireless networks. *Telecommunication Systems Journal 3* (2001), 63–84.

[42] ZHENG, R., HE, G., GUPTA, I., AND SHA, L. Time idexing in sensor networks. In *1st IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS)* (2004).