On the Selection of Cluster Heads in MANETs

Article *in* International Journal of Computer Science Issues · March 2012

4 authors:

Dang Nguyen
Torrens University Australia

Pascale Minet
National Institute for Research in Computer Science and Control

T. Kunz
Carleton University

L. Lamont
Communications Research Centre Canada

On the Selection of Cluster Heads in MANETs

Dang Nguyen[1], Pascale Minet[2], Thomas Kunz[3] and Louise Lamont[1]

[1] Communications Research Centre
Ottawa, ON K2H 8S2, Canada

[2] INRIA Rocquencourt
Rocquencourt, Le Chesnay Cedex 78153, France

[3] Dept. of Systems and Computer Engineering, Carleton University
Ottawa, ON K1S 5B6, Canada

Abstract

Clustering schemes offer a practical way of providing scalability when dealing with large and dense Mobile Ad hoc Networks (MANETs). The feasibility of a clustering method can be primarily determined by the complexity of the cluster head selection. Optimizing the cluster head selection allows for the network to be more efficient by minimizing the signaling overhead while ensuring that the network connectivity is maintained despite topology changes. In this paper, we investigate the problems of cluster head selection for large and dense MANETs. Two variants of the cluster head selection are examined: (1) the *distance-constrained selection* where every node in the network must be located within a certain distance to the nearest cluster head; and (2) the *size-constrained selection* where each cluster is only allowed to have a limited number of members. We show that the problem of minimizing the set of cluster heads is NP-hard for both variants. We propose two distributed selection algorithms, each having logarithmic approximation ratio, for these variants. We also discuss, using simulations, the resulting cluster size distribution and cluster head density, which impact the efficient operation of the network.

Keywords: *MANET, scalability, clustering algorithms, complexity, NP-complete.*

1. Introduction

In the near future, the US military's Joint Tactical Radio System (JTRS) [1] is expected to create radios that work together to form autonomous *ad hoc* networks. Also, the US DARPA's Wireless Network after Next program (WNaN) [2, 3] aims at developing technologies and concepts enabling the deployment of massively dense networks. The technology created by the WNaN program is expected to provide reliable and highly-available battlefield communication systems at low operating cost. As a result, there will be challenges for routing protocols to support distributed and adaptive network operations in these large, dense and scalable MANETs.

Out of many existing MANET routing protocols, OLSR [4] is being considered as a very potential candidate for IETF standardization and for military networking deployment. OLSR is a proactive protocol, which means the node knowledge about the network topology is periodically refreshed.

When the size of the network grows, the amount of signaling overhead also increases to maintain the topology updates. One of the main issues of a MANET's routing protocol is hence its capacity to scale on large and dense networks. The two most popular techniques to reduce signaling overhead in MANETs are Fish Eye [5, 6] and clustering [7].

With the Fish Eye technique, the frequency of topology updates is inversely proportional to the distance to the updating source. Instead of sending signaling messages to distant nodes at the same rate as to nearby nodes, Fish Eye modifies the routing protocols such that these messages are only forwarded at a lower rate beyond some distance thresholds. A strong advantage of the Fish Eye technique is that the routing protocols can easily be modified to enable Fish Eye capability in practical implementation. Also, Fish Eye routers do not need extra network interfaces to relay information as compared to cluster heads' requirement in some cluster-based techniques. However, the Fish Eye technique still keeps a flat network architecture. Thus, every node still relays signaling messages for every other node, less frequently though.

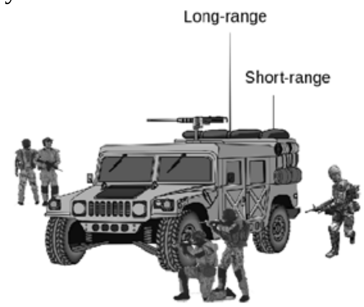
In cluster-based routing, the network is divided into clusters. Each cluster has a cluster head (CH) node and some ordinary member nodes. MANET routing protocols are run in each cluster and their signaling messages are to propagate only within the cluster. The CHs notify each other about their cluster's members frequently using a different communication channel. Inter-cluster communications are relayed by CHs. The CHs may in turn form another MANET and be clusterized to an upper level if needed.

In order to reduce the overhead of the CH communications, the number of clusters must be minimized in the whole network. The CHs are thus spaced out to cover all nodes of the network and this also improves the spatial reuse of CH intra-communications. Therefore, most cluster-based techniques form non-overlapping clusters where CHs have multiple network interfaces with different communication ranges (*e.g.*: short range for intra-cluster and long-range for inter-cluster communications.) Notice that cluster-based technique can also be applied to MANETs where the nodes only have single network interface. In this situation, the inter-communication between distant CHs takes place as point-to-point communications. The traffic is then relayed by ordinary, intermediate nodes sitting between these CHs. While the network's communication performance can be different depending on the number of wireless interfaces each node has, the problem of CH selection is fundamentally unchanged.

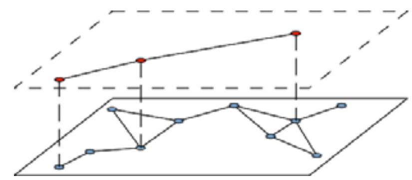
Compared to the flat network architecture inherent to the Fish Eye technique, the hierarchical structure of cluster-based routing is more suitable for a well-defined, multilevel tactical military network. In practice, the *Hierarchical OLSR* protocol (HOLSR [8]) has implemented a cluster-based routing mechanism for tactical MANETs. Figure 1(a) illustrates a combat unit in a tactical MANET. The combat unit includes a vehicle and the ground troops assigned to it. Communications between troops of different units are relayed by their vehicles. Each vehicle has two radio interfaces: short- and long-range. The short-range interface allows the vehicle to communicate with neighboring vehicles, with distances from hundreds meters to one kilometer. The long-range interface allows for communication with other vehicles farther than several kilometers. Each combat unit is represented by a node of the graph in Figure 1(b). The short-range radio interface allows units to form a multihop MANET.

With cluster-based routing, three CHs are selected among the nodes in Figure 1(b). Each CH covers a cluster encompassing its direct neighbors. The CHs then communicate with each other using the long-range interfaces. Therefore, they may form another multihop MANET. The communications between nodes from different clusters are relayed by CHs.

The CH selection is static in the current implementation of HOLSR. The CHs are chosen before the network's deployment. They broadcast messages inviting other nodes to join their clusters as a function of the nodes' distance to the nearest CH. No new CH is selected during the network's operation. This static selection may lead to problems of CHs' availability due to node mobility or due to CHs' failure.



(a) A combat unit (equivalent to an HOLSR node).



(b) Three CHs selected to cover all nodes.

Fig. 1. Selection of CHs in a tactical MANET.

In this paper we investigate the selection of CHs in a distributed environment such as MANET. We derive new results on the complexity and efficiency of two variants of the CH selection: distance-constrained and size-constrained clustering. The analysis of our simulations allows for the recognition of some properties that are most relevant to the overhead and the performance of these networks.

The rest of this paper is organized as follows. We present in Section 2 some existing work on CH selection in MANETs. The complexity of CH selection is investigated in detail in Section 3. We also present distributed algorithms to select CHs in MANETs. We analyze in Section 4 some properties of these algorithms, obtained by simulations, that are most relevant to the overhead and the efficiency of the network. Finally, we conclude this paper in Section 5.

2. Related Work

CH selection has extensively been studied in the literature of wireless *ad hoc* networks. It was showed in [9, 10] that using clusters for data-aggregation in large-scale sensor networks can significantly improve the sensors' lifetime. In [9], Heinzelman *et al* propose a protocol (LEACH) that allows nodes to select CHs using a distributed algorithm. Each sensor takes its turn as CH so that their energy consumption is balanced. LEACH ensures that the network has on average a fixed, predefined number of CHs at any time. Chen *et al* [10] improve this approach by first estimating the optimal number of clusters to efficiently utilize data correlation of sensors. A new random CH selection algorithm is then proposed, aiming at minimizing the distance between the CHs and their members.

Koshy *et al* [11] show that information entropy (used by the authors as a way to classify nodes as conservative or exploratory based on their activities) can also serve as a metric to form clusters. Nodes with low level of activities are more likely to become CHs. This method may therefore produce stable CHs.

In [12], Xia *et al* propose a distributed CH selection protocol that forms clusters of nodes having similar sensed data in order to optimize the data aggregation at the CHs. Their protocol also considers including into the cluster the nodes located at any distance up to h -hops away from the CH. Thus, their work is closely related to our distance-constrained CH selection with an additional constraint on the node's data similarity.

Regarding MANETs, Chinara *et al* report in [7] an interesting survey on clustering algorithms, ranging from nodes' ID-based selection to mobility and connectivity metric-based selection. They show that while ID-based selection produces a fast and stable cluster setup, it suffers from the rigidness of the CHs' structure, because the same nodes are often selected independently of the network topology. Topology-dependent CH selection (based on mobility and connectivity metrics) can produce a more evenly distributed CH set. However, they may require a larger cluster setup time. An example of clustering based on mobility consideration is given in [13] by Konstantopoulos *et al*.

We choose to consider the CH selection in this paper uniquely with the constraints related to the network topology graph, *i.e.* limiting the distance (in number of hops) between each CH and its members or limiting the size of each cluster. The reason behind those limitations is because other metrics (e.g.: energy, traffic load, mobility factors) can often be modeled using an appropriate weighted graph topology. For example: energy-saving CH selection in a network can be modeled by a CH selection in a weighted graph (the weight of each node is inversely proportional to its remaining energy amount) in which the sum of all CHs' weight is minimized. More generally, in graph theory, CH selection is studied with the *dominating set* problems [14].

Out of the two variants of CH selection that we present in this paper, the distance-constrained CH selection is cited as a known variant of the dominating set problem in [14]. Amis *et al* present in [15] a proof showing the NP-completeness of this variant. The authors also propose a heuristic to select CHs based on the nodes' ID. However, this heuristic is known to fail to provide a good solution in some pathological cases, for example: when the nodes' ID are monotonically increasing or decreasing in a straight line. Also, the efficiency of the proposed heuristic, represented by the approximation factor of its result compared to an optimal solution, has not been investigated.

The second variant that we examine in this paper, the size-constrained CH selection, is more related to the work done by Nam *et al* [16] where the CH selection tries to form clusters of equal size. Chatterjee *et al* [17] also propose a distributed clustering algorithm that takes into consideration various parameters such as ideal degree, transmission power, mobility, *etc.*, while limiting the number of members in each cluster.

To the best of our knowledge, there is no known result on the complexity of the size-constrained CH selection. Also, the existing work done on CH selection does not investigate the approximation factor of the proposed solutions for both problems that we examine.

We present in this paper a proof showing the NP-completeness of the size-constrained CH selection. A new proof, which is significantly shorter than the one in [15], is also presented for the distance-constrained CH selection problem. Moreover, we propose a distributed algorithm for each problem and show that they can achieve logarithmic approximation factors, which is known (see Feige [19]) to be best possible unless NP has super-polynomial time algorithms.

Notice that other variants of CH selection exist. For example, Kuhn *et al* [18] propose two algorithms of CH selection such that each node is a member of no less than k different clusters to ensure fault tolerance. The study of these variants is beyond the scope of this paper because we are primarily concerned with the minimization of the number of CHs in the whole network. However, we acknowledge in our analysis that the number of clusters that encompass a node is a factor reflecting the robustness of the clustering scheme.

Given the *ad hoc* nature of MANET routers and their low computational capacity, it is thus important that we investigate the complexity of these CH selection variants and propose distributed algorithms that can be applied to a tactical MANET environment.

3. Cluster Head Selection

We study in this section two variants of the CH selection for MANETs. The first variant selects CHs such that every dependant node is within a distance h hops from the nearest CH. The second variant selects CHs such that the size of each cluster is not larger than τ . We will discuss the complexity of both problems and derive distributed CH selection algorithms that are applicable to MANETs.

In addition, a third variant, called *distance-and-size-constrained CH selection*, which is a combination of the two variants described above, is also examined.

3.1 Distance-constrained CH Selection

In this section we consider the selection of CHs in a MANET of n nodes such that every node in this network is within distance h hops of a CH, for a given positive h. Such a set of CHs is said to *cover within h hops* the whole network. It is natural to seek the minimum set of CHs to reduce the communication overhead between CHs.

To start, we state a result on the NP-completeness of the decision problem of finding such a set of size no larger than k CHs. Then, we present a greedy distributed algorithm allowing to select the CHs with an approximation factor of $\min(\ln\Delta^h, \ln n)$, where Δ is the maximum degree of the topology graph.

3.1.1 Complexity

Let $G=(V,E)$ be a graph representing the network topology, $|V|=n$. Each vertex $v\in V$ represents a node and for all vertices $u,v\in V$, $(u,v)\in E$ if and only if two nodes u and v are direct neighbors. Let h be a positive number, the minimum CH set of the MANET is then represented by the minimum set of vertices $S\subseteq V$ such that for every vertex u, either $u\in S$ or there exists a vertex $v\in S$ such that $d(u,v)\le h$. $d(u,v)$ denotes the shortest distance between nodes u and v in terms of hop. Such a set S is called *distance-h dominating set* of G (cf. [14]).

Notice that if $h=1$ then the problem of finding such a minimum set S is identical to the *minimum dominating set* problem, which is equivalent to the NP-hard *minimum set cover* in [20].

We examine the decision problem of the distance-h dominating set, defined as follows. Let $k<n$ positive, does the network admit a set of CHs of size at most k such that each node is either a CH or is within distance h hops away from a CH? One can see that such a set exists if and only if G admits a distance-h dominating set of size at most k.

Theorem 1: *The decision problem of the distance-h dominating set is NP-complete.*

Proof: It is easy to verify that this problem is in NP. Given a set S, $|S|\le k$, it can be checked in polynomial time that every vertex of G is either in S or within distance h to a vertex in S by calculating the shortest path from all vertices in S to all vertices in $V\setminus S$.

To prove NP-completeness, we use induction on h by reducing the problem distance-$(h-1)$ dominating set to the problem distance-h dominating set. Notice that this problem is known to be NP-complete when $h=1$.

Let $G=(V,E)$ be a graph. We construct a new graph G' by extending G in the following manner: for each $v\in V$, we add a new vertex v' and an edge connecting v and v'. Formally, $G'=(V',E')$ where $V'=V\cup\{v' \mid v\in V\}$ and $E'=E\cup\{(v,v') \mid \forall v\in V\}$. This construction is polynomial time. Our goal is to show that G has a distance-$(h-1)$ dominating set of size at most k if and only if G' has a distance-h dominating set of size at most k.

Let S be a distance-$(h-1)$ dominating set of G of size k. It is clear that S is also a distance-h dominating set of G'. Because for each vertex $v\in V'\setminus S$, if $v\in V$ then there is $s\in S$ such that $d(s,v)\le h-1<h$ by the definition of S. If $v\in V'\setminus V$ then v is connected to a vertex $v^*\in V$. Again, there exists $s\in S$ such that $d(s,v^*)\le h-1$ in G leading to $d(s,v)\le h$ in G'.

Now, let us assume that S' is a distance-h dominating set of G' of size k. We construct a set S from S' as follows. $S=(S'\cap V)\cup\{s\in V \mid s'\in S'\setminus V\}$. We have $|S|\le|S'|=k$ by construction and S only contains vertices from V. For each $v\in V\setminus S$ let $s\in S$ such that $d(s,v)$ is minimum. It is impossible that $d(s,v)\ge h$ in G since it would imply $d(s',v')\ge h+1$ with v' the extended vertex of v in G' and for all $s'\in S'$, a contradiction of the definition of S'. Therefore, S is a distance-$(h-1)$ dominating set of G of size at most k.

3.1.2 Greedy distributed algorithm for CH selection

There are many centralized algorithms to approximate the minimum dominating set (cf. [21, 22, 23]). However, it is known by Feige [19] that the minimum dominating set cannot be approximated within a ratio of $(1-\varepsilon)\ln n$, for any $\varepsilon>0$, unless NP has $n^{O(\log\log n)}$ time algorithms. Therefore, known polynomial time approximation algorithms for this problem, which produce an approximation factor of $\ln n$, are essentially best possible. If the maximum degree of the network graph is Δ, then an approximation factor of $\ln \Delta$ can be achieved.

For a given positive h, we design a greedy, distributed algorithm appoximating the minimum dominating set for the selection of a distance-h dominating set. Let v be a node, the distance-h neighborhood of v, denoted as $N_h(v)$, contains all nodes within h hops from v. The distance-h degree of node v is $d_h(v) = |N_h(v)|$. Let $W_h(v)$ be the set of uncovered nodes in $N_h(v)$ and $w_h(v) = |W_h(v)|$. We assume there exists a distance-h neighborhood discovery protocol that allows each node v to know $N_h(v)$, $W_h(v)$ and $w_h(u)$ for all $u \in N_h(v)$. Typically, for $h = 2$, the NHDP protocol for MANETs by Clausen *et al* [24] can easily be adapted to satisfy this requirement.

Each node v executes the following greedy algorithm to select the CHs according to the distance-h constraint:

Algorithm

1. While v is still uncovered:
2. If there is $u \in W_h(v)$, $u \neq v$, such that $w_h(u) = \max(w_h(z) \mid z \in W_h(v))$ then send a message to $W_h(v)$ declaring the wish to select u as CH. In case of a tie, then choose the node having the largest ID.
3. If all nodes in $W_h(v)$ select v as CH, then v sends a message to $W_h(v)$ to announce it is becoming a CH. v is marked as covered.
4. If v has sent a message to select u as CH and has received a message from u announcing that it becomes a CH then v is marked as covered.
5. end while.

In this greedy algorithm, at least one CH is selected after each round of its execution. To see that this is true: it is true for the first execution round in which there is at least one node u selected by all its h-hop neighbors (at least the node u with the largest $w_h(u)$ in the whole network will be selected.) Node u then forms its cluster and this cluster is removed from the topology graph because the cluster's nodes are marked as covered. The algorithm is re-executed with this new topology graph.

Therefore, the time complexity of the CH selection is at most linear in the size of the network. We also know the approximation factor achievable by this algorithm based on a similar result on the greedy set-covering algorithm in [21] (see also Chvátal [22]), to which the interested reader may refer for full details.

Theorem 2: *This greedy algorithm yields a dominating set of size of* $O(|S^*| \min(\ln \Delta^h, \ln n))$, *where* S^* *is the minimum distance-*h *dominating set for the given instance.*

Proof: It is known from [21] that the greedy algorithm for the dominating set problem can achieve an H_Δ-approximation. H_i is the i th harmonic number. If we expand the notion of neighborhood to consider all neighbors within h hops, then the approximation factor becomes $H_{\Delta[h]}$ where $\Delta[h] = \max_{v \in V}(|N_h(v)|) \le n$.

On the other hand, we have: $\forall v \in V$,

$$|N_h(v)| \le 1 + \Delta \sum_{i=0}^{h-1} (\Delta - 1)^i = 1 + \Delta \frac{(\Delta-1)^h - 1}{\Delta - 2}.$$

Thus, $\Delta[h] \le \min(\Delta^h, n)$ and

$$H_{\Delta[h]} = \sum_{i=1}^{\Delta[h]} \frac{1}{i} \le \min(\ln \Delta^h, \ln n) + O(1).$$

3.2 Size-constrained CH Selection

There is a major drawback with the previous selection of a CH set. Because this mode of selection is based solely on the distance constraint, it offers no control over the size of each cluster. If some clusters are too large and the CHs have to relay a high amount of control traffic for their dependants then congestions may occur in the network. It can directly impact the network's quality of service.

Figure 2 shows the distribution of the cluster size for a network of $n = 100$ nodes with node density $\nu = 20$. This distribution is obtained by averaging the simulation results of 20000 random network scenarios. The CHs are selected according to the distance-2 constraint, *i.e.* each node is either a CH or is within 2 hops from a CH. The x-axis depicts the size of clusters and the y-axis the percentage of nodes being in a cluster of that size. This percentage is calculated over 100 nodes and over 20000 random scenarios that we simulated. We can observe that the cluster size's distribution is highly uneven: 62% of nodes in the network are in clusters that have more than 60 dependants, whereas 29% of nodes are in clusters that only have 20 dependants or less. With the majority of nodes being dependants of large clusters, the network traffic can be congested due to the bottlenecks at the CHs.

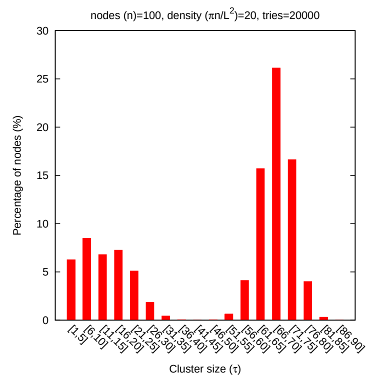


Fig. 2. Distribution of cluster size for 100 nodes over 20000 scenarios.

Therefore, we consider a second method of CH selection for MANETs, called *size-constrained CH selection*. Each CH is only allowed to include a maximum number τ of nodes in its cluster. If some nodes in its neighborhood are still uncovered, then a new CH must be selected to cover them. We call this problem *size-* τ *dominating set*.

Our goal is to seek a minimum set of CHs to reduce the overhead between them. Similar to the distance- h dominating set problem, we find that the decision problem for the size- τ dominating set is also NP-complete for a general graph and $\tau..2$. We then present a distributed algorithm to select a small set of CHs with size constraint τ.

3.2.1 Complexity

The decision problem of the size- τ dominating set is defined as follows. Let $k < n$ positive, can the network be partitioned into at most k clusters, each CH has no more than τ dependants and is at distance 1-hop from its dependants? It is trivial that such a partition exists if and only if the network graph can be partitioned into at most k subgraphs, each isomorphic to a star of degree at most τ.

Notice that if $\tau = 1$ then the set of clusters becomes a maximum matching of the graph, which implies the problem can be solved in polynomial time using Edmonds' algorithm [25] for any graph. On the other hand, if $\tau..\Delta$, then it is equivalent to the classical NP-hard *minimum dominating set* problem.

Theorem 3: *The decision problem of size-* τ *dominating set is NP-complete for all* $\tau..2$.

Proof: This problem is clearly in NP since it can be verified in polynomial time that each subgraph in a set of k subgraphs is isomorphic to a star of degree at most τ.

To show that it is NP-complete, we reduce a general instance of the minimum dominating set problem to our problem. Notice that we only need to prove NP-completeness for all graphs having $\Delta = \tau + \lambda$, for any fixed $\lambda..1$.

Let $G_0 = (V_0, E_0)$ be a graph with maximum degree $\Delta_0,, \tau$. We construct a graph G from G_0 as follows. First, we create $(\tau - \Delta_0 + \lambda)$ copies of G_0, denoted as $G_i = (V_i, E_i)$ for $i = 1..(\tau - \Delta_0 + \lambda)$. Each G_i is isomorphic to G_0. Let $v_0 \in V_0$ be a node having maximum degree in G_0 (*i.e.* $d(v_0) = \Delta_0$) and $v_i \in V_i$ be the copy of v_0 in G_i. We obtain G by connecting all v_i together, for $i = 0..(\tau - \Delta_0 + \lambda)$. This construction is polynomial time. Figure 3 shows the construction of G from an example of G_0.

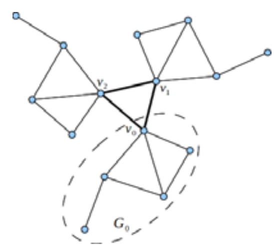


Fig. 3. Construction of the graph G from G_0 with $\tau = 4$ and $\lambda = 1$.

By the construction of G, its maximum degree is $\Delta = d_G(v_i) = \tau + \lambda$. Our goal is to show that G_0 admits a dominating set of size at most k if and only if G admits a size- τ dominating set of size at most $k(\tau - \Delta_0 + \lambda + 1)$.

Let $S_0 \subseteq V_0$ be a dominating set in G_0 of size k. Let $S_i \subseteq V_i$ be the copy of S_0 in graph G_i. It is trivial that $S = \cup_{i=0}^{\tau-\Delta_0+\lambda} S_i$ is a dominating set of G. Because each $s \in S_i$ covers a maximum number of $\Delta_0,, \tau$ in G_i, S is also a size- τ dominating set of G. We also have $|S| = \sum_{i=0}^{\tau-\Delta_0+\lambda} |S_i| = k(\tau - \Delta_0 + \lambda + 1)$.

We assume now that G admits a size- τ dominating set of size $k(\tau - \Delta_0 + \lambda + 1)$, denoted as S. Let $S_i = S \cap V_i$. Because V_i and V_j are disjoint for all

$i \neq j$, we have $|S| = \sum_{i=0}^{\tau-\Delta_0+\lambda} |S_i|$ and $S_i \cap S_j = \varnothing$. Two cases are possible:

1. $\forall i, |S_i| = k$. If $v_0 \in S_0$ then after removing the edges connecting v_0 to $v_i \in V_i$, we obtain the graph G_0 with a dominating set S_0 of size k. Otherwise, if there is j such that $v_j \in S_j$ then we can obtain by the same process a graph G_j isomorphic to G_0 with a dominating set S_j. Now, if $\forall i, v_i \notin S_i$ then by the construction of G every v_i is covered by a vertex $u_i \in S_i \subseteq V_i$ (*i.e.* none of the v_i is covered by another v_j.) This means removing the edges connecting all v_i together does not change the coverage in G. Hence, S_0 is a dominating set in G_0.

2. $\exists i \in \{0, \ldots, \tau - \Delta_0 + \lambda\}$, such that $|S_i| < k$. It is straightforward in this case that $S_i \cup \{v_i\}$ is a dominating set in G_i isomorphic to G_0, of size at most k.

3.2.2 Greedy distributed algorithm for CH selection

Similarly to the selection of CHs based on the distance constraint, we design a greedy, distributed polynomial time algorithm for our problem of size-constrained CH selection.

We use the same notations as with the distance-constrained CH selection. In particular, $N(v)$ denotes the 1-hop neighborhood of v and $W(v) \subseteq N(v)$ is the set of v's uncovered neighbors. Let $w(v) = |W(v)|$. We assume there is a neighborhood discovery protocol that allows each node v to know $N(v)$, $W(v)$ and $w(u)$ for all $u \in N(v)$.

This algorithm is executed by v until v is covered, *i.e.* v becomes a CH or a CH's dependant.

Algorithm

1. While v is still uncovered:
2. If there exists $u \in W(v)$, $u \neq v$, such that $w(u) = \max(w(z) \mid z \in W(v))$ then send a message to $W(v)$ declaring the wish to select u as CH. In case of a tie, then choose the node having the largest ID.
3. If all nodes in $W(v)$ wish to select v as CH, then v sends a message to $W(v)$ to announce it is becoming a CH. This messsage contains a list of at most τ neighbors that v has chosen to include into its cluster. These neighbors u are chosen according to the increasing order of their $w(u)$. v is marked as covered.
4. If v has sent a message to select u as CH and has received a message from u announcing that it becomes a CH with v in the list of neighbors selected by u then v is marked as covered. v becomes u's dependant.
5. end while.

The time complexity of this greedy algorithm is at most linear in the size of the network because at least one new CH is selected after each round of execution of the algorithm. The following theorem allows for calculating the approximation factor of the solution.

Theorem 4: *This algorithm computes a* $\ln \tau$ *-approximation compared to the optimal size-* τ *dominating set.*

Proof: Let v be a CH selected in an optimal size-τ dominating set. Each dependant of v is given an equal cost of $\frac{1}{\min(d(v),\tau)}$. We can calculate the total cost of $N(v)$, denoted by $C^*(v)$ as follows. $C^*(v) = 1$ if $d(v) < \tau$; and $C^*(v) = \frac{d(v)}{\tau}$ if $d(v) \geq \tau$. On the other hand, the total cost of $N(v)$ in a size-τ dominating set obtained by the greedy algorithm is $C(v)$, calculated as follows.

If $d(v) < \tau$ then $C(v) \leq \sum_{i=1}^{d(v)+1} \frac{1}{i} \leq \ln(d(v)) + O(1)$. Thus, $\frac{C(v)}{C^*(v)} \leq \ln \tau + O(1)$.

If $d(v) \geq \tau$ then for each node $z \in W(v)$ becoming a member of another cluster before v is selected as CH, a cost of c_z is assigned to this node: $c_z = \frac{1}{\tau}$ if $|W(v)| \geq \tau$, and $c_z \leq \frac{1}{|W(v)|}$ otherwise. The worst case occurs when there are $d(v) - \tau$ such nodes in $N(v)$ becoming members of other clusters with each having a cost of $\frac{1}{\tau}$; and for the τ remaining nodes, they become members of other clusters with their respective costs of $\frac{1}{\tau}, \frac{1}{\tau-1}, \ldots, \frac{1}{1}$. Therefore, $C(v) = \sum_{z \in N(v)} c_z \leq \frac{d(v)-\tau}{\tau} + \sum_{i=1}^{\tau} \frac{1}{i}$. We then have, with H_τ being the τ th harmonic number: $\frac{C(v)}{C^*(v)} \leq \frac{\tau}{d(v)}\left(H_\tau - 1\right) + 1 \leq \ln \tau + \mathrm{O}(1)$.

3.3 Distance-and-size-constrained CH Selection

The size-constrained CH selection problem and algorithm presented in the previous section work explicitly with 1-hop neighbors only. We examine in this section a third variant of the CH selection, which is a combination of the two previous variants, called *distance-and-size-constrained CH selection*. The corresponding decision problem, called *(distance-h, size-τ)–dominating set*, is defined as follows. Let $k < n$ positive, can the network be partitioned into at most k clusters such that each CH has no more than τ dependants and is at distance at most h hops from its dependants?

Analogously, we find that this problem is also NP-complete for all $h \geq 1$ and for all $\tau \geq 2$. Due to space limitations, we only present the proof for the NP-completeness of this problem. The size-constrained CH selection algorithm can be easily adapted to the distance-and-size-constrained CH selection and will be left to the interested readers.

Theorem 5: *The decision problem of the (distance-h, size-τ)–dominating set is NP-complete for all $h \geq 1$ and $\tau \geq 2$.*

Proof: This proof is similar to the one of Theorem 3. Starting from the premise that the distance-h dominating set is NP-complete (Theorem 1), we reduce a general instance of the distance-h dominating set to the (distance-h, size-τ)-dominating set problem. We present here the sketch of this proof.

To start with, this problem is clearly in NP since it can be verified in polynomial time if each CH has at most τ dependants and is at distance h hops, at most, from its dependants.

By a similar construction of a graph as in Theorem 3, let $G_0 = (V_0, E_0)$ be a graph with $\Delta_0[h] = \max_{v \in V_0}(|N_h(v)|) \leq \tau$. We create $(\tau - \Delta_0[h] + \lambda)$ copies of G_0 and connect them together to obtain the graph $G = (V, E)$, with $\Delta[h] = \max_{v \in V}(|N_h(v)|) \geq \tau + \lambda$.

By considering two cases as done in the proof of Theorem 3, it can be showed that G_0 admits a distance-h dominating set of size k if and only if G admits a (distance-h, size-τ)--dominating set of size $k(\tau - \Delta[h] + \lambda + 1)$. This completes the proof.

4. Simulation

We present some simulation results in this section. Our simulations aim at showing, for the distance-constrained and the distance-and-size-constrained CH selection, the parameters that influence the network overhead such as the total number of clusters in the network and the CH density (*i.e.* the average number of CHs in the distance-h neighborhood of each node.)

In the following simulations, the distance-h CH selection algorithm is executed with $h = 2$. The size-τ CH selection algorithm is executed with various values of τ and also extended to cover the 2-hop neighbors. That means a CH will: (1) include a maximum of τ nodes among its uncovered direct neighbors to its cluster and (2) if all direct neighbors are covered and there is still room then include two-hop neighbors (only the ones reachable through a direct neighbor already in the cluster) until arriving at τ dependants.

It is worth pointing out that while it is feasible to implement the CH selection algorithms to cover $h \geq 3$ hops neighborhood, there may be an overhead tradeoff to consider. Such an implementation requires a signaling protocol to collect information from all nodes up to h-hops neighborhood (an example of this implementation is to retransmit NHDP's Hello messages up to h hops.) This can lead to a significant increase in local signaling overhead as the number of Hello messages in the h-hops neighborhood grows in $\mathrm{O}(\Delta^h)$.

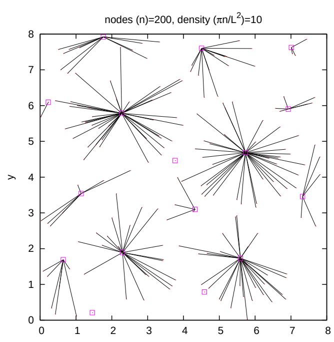
Our simulator is written in the C language. We assume there is no loss at the communication level. In a typical simulation, our program generates a random network topology according to some input parameters. Then the CH selection algorithms are executed by the nodes on this network topology and the parameters of interest are reported. The input parameters are the total number of nodes n in the network, the average node density ν and, only for the size-constraint algorithm, the maximum allowable size τ of each cluster. For a particular simulation configuration (*i.e.* a particular set of input parameters), the algorithms are executed on 20000 randomly generated network topologies and the results are averaged.

To generate a network topology from the parameters n and ν, we assume that the communication range of each node is unitary. Therefore, two nodes are direct neighbors if and only if their euclidean distance is no more than 1. The n nodes are then randomly placed on a square of size L^2 with $L = \sqrt{\frac{\pi n}{\nu}}$.

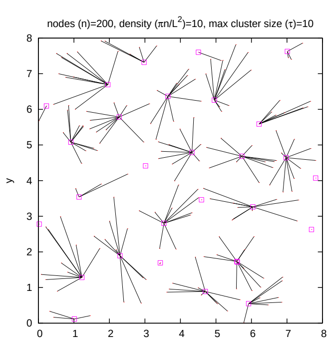
To start, we compare on an identical network topology the selections of CHs with and without size constraint. Then we continue by presenting the parameters of interest for each algorithm.

4.1 Comparison of Cluster Formation

Figure 4 compares the selections of CHs with and without size constraint for the same network topology. This network has $n = 200$ nodes with an average density of $\nu = 10$ nodes per unitary disc.



(a) No cluster size constraint.



(b) Maximum cluster size $\tau = 10$.

Fig. 4. Comparison between size-constrained and no size-constrained for distance-2 CH selection in a network of 200 nodes.

Figure 4(a) shows the formation of clusters when no size constraint is specified. We can see that the distribution of the cluster sizes is highly uneven between the clusters: four clusters (25% of the total clusters) have a size more than double that of the other clusters. Thus, the CHs of those large clusters may encounter traffic congestion. It is worth noting that, because of the border effects, nodes located far from the network borders usually have more neighbors, thus they have higher degrees, than nodes in the border's vicinity. This effect implies that the nodes far from the borders are susceptible to be among the first selected CHs by this algorithm.

On the other hand, Figure 4(b) shows the cluster formation with a maximum size constraint $\tau = 10$, which equals the average node density. The cluster formation is more regular with most clusters having similar size. Thus, the network load is distributed more evenly among the selected CHs. However, there are almost twice as many CHs selected as in Figure 4(a) (29 CHs versus 16 CHs.)

4.2 Number of Clusters and CH Density

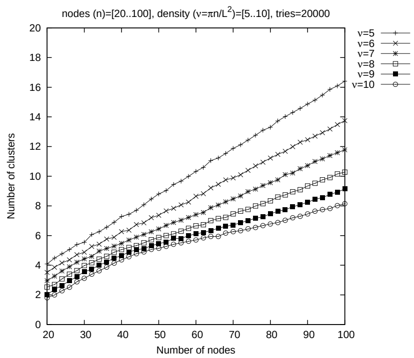
We examine in this section two parameters of interest that can influence the network efficiency: the number of clusters and the CH density. The number of clusters indicates the overhead of the network at the CH level. The CH density is calculated as the average number of CHs that each node can find in its 2-hop neighborhood. Therefore, the CH density reflects the robustness of the CH selection algorithm: in case of a CH failure, its dependants may backup immediately to an existing CH found in the 2-hop neighborhood.

It is worth mentioning that we may need a protocol to support the recovery from a CH failure. Such a protocol would allow nodes having a failed CH to know who the alternative CHs are, and negociate with candidate CHs to find a cluster to join. We may also need to allow a temporary moment when the backup CHs have to accept new nodes even if their size exceeds the limit τ before a new CH selection procedure is triggered. Nevertheless, having multiple CHs already selected in the neighborhood, and under the assumptions that the CHs exchange their database of members with their CH peers with regard to an eventual backup, can help to recover more quickly from a CH failure compared with having to re-elect a new CH and waiting for this CH to collect all the information about the members before disseminate it into the networks.

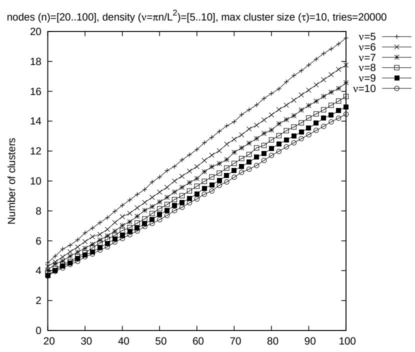
4.2.1 Number of clusters

Figure 5 compares the number of clusters formed by each selection algorithm. The total number of nodes in the network varies from 20 to 100 nodes. The node density varies from 5 to 10 nodes per unitary disc. For size-constrained CH selection (Figure 5(b)), we fixed $\tau = 10$.

We can see that the number of clusters formed by both algorithms increases almost linearly with the number of nodes in the network. This trend is true independently of the node density. In sparse networks ($\nu = 5, 6, 7$), there are slightly more clusters when the cluster size is limited to $\tau = 10$ than when it is not. This gap becomes larger for dense networks ($\nu = 10$): 14 clusters with size constraint compared to 8 clusters without size constraint for a network of 100 nodes. However, it is still a very efficient way to reduce signaling overhead compared to a flat network, because the number of CHs is less then 20% of the total number of nodes.



(a) No cluster size constraint.

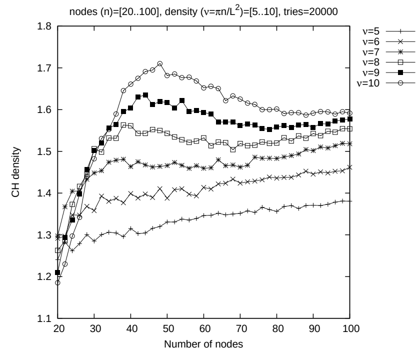


(b) Maximum cluster size $\tau = 10$.

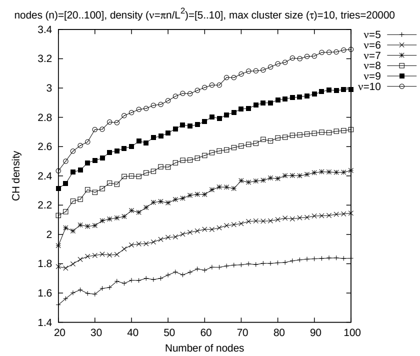
Fig. 5. Comparison of the number of clusters between size-constrained and no size-constrained for distance-2 CH selection in a network of $n = 20..100$ nodes and density $\nu = 5..10$.

4.2.2 CH density

Figure 6 compares the CH density between the two selection algorithms, with the same network configurations as above. We can see that for both algorithms, the CH density increases very slowly with the number of nodes.



(a) No cluster size constraint.



(b) Maximum cluster size $\tau = 10$.

Fig. 6. Comparison of the CH density between size-constrained and no size-constrained for distance-2 CH selection in a network of $n = 20..100$ nodes and density $\nu = 5..10$.

Notice that for dense networks ($\nu = 8, 9, 10$), the curves in Figure 6(a) show a slight decrease in CH density for $n = 40..80$ nodes. This is due to the border effects: for dense and small networks ($\nu = 8, 9, 10$ and $n \leq 40$ nodes), there are more nodes affected by the border effects. Therefore, some CHs located in the border areas cover less nodes than other CHs in the center, resulting in a higher ratio of CHs per node.

It is clear from Figure 6(a) that the distance-2 CH selection without cluster size constraint cannot ensure backup in the event of CHs' failure, as its average CH density is only 1.7 CHs per node at most. That means if some CHs fail then at least 30% of the dependants cannot find a second CH in their 2-hop neighborhood for an immediate backup. With the size constraint CH selection (cf. Figure 6(b)), most networks with density $\nu \geq 7$ can ensure an immediate backup since the CH density is always higher than 2.

5. Conclusions

We investigate in this paper the complexity and performance of different cluster head (CH) selections in MANETs. Two variants of CH selection are examined. The first variant (*a.k.a.* distance-constrained) selects a set of CHs such that every node in the network is either a CH or is located within distance h hops away from the nearest CH. The second variant (*a.k.a.* size-constrained) limits the maximum size of each cluster to τ members. A third variant, combining the distance and size constraints, is also presented.

The decision problems of these variants are showed to be NP-complete for a general network graph. We propose two distributed algorithms for these CH selections. Each algorithm has logarithmic approximation ratio, which is known to be best possible unless NP has superpolynomial time algorithms. The time complexity of these algorithms is at most linear in the size of the network.

Our simulation results show that the distance-constrained CH selection can find a smaller CH set compared to the distance-and-size-constrained selection. However, the cluster size is unevenly distributed among the clusters. This may create congestion at some CHs if they have to relay a large amount of traffic for their dependants. The simulations also show that the distance-and-size-constrained CH selection can solve this issue by selecting more CHs in the network. The clusters then have similar size.

Also according to our simulations, while the number of clusters in the network increases linearly with the network size for both algorithms, the CH selection with size constraint can offer a more robust connectivity to the dependants. Its CH density is higher than 2 for most network configurations. That means if some CHs fail, their dependants may be able to find an existing CH in the neighborhood ready for a quick backup. Notice that this backup feature needs an additional protocol to help nodes recovering from a CH failure, which is a subject for further research.

Another issue relevant to the clustering performance is the management of node mobility and topology changes. We believe that the consideration of topology changes in CH selection algorithms is challenging and has the merit of being examined separately in a future study.

Acknowledgments

This work is funded by Defence Research & Development Canada (DRDC).

References

[1] A. Feickert, “The joint tactical radio system (jtrs) and the army’s future combat system (fcs): Issues for congress,” UNT Digital Library, Washington D.C., USA, November 2005.

[2] DARPA, “Wireless network after next,” Proposer Information Pamphlet, http://www.darpa.mil/sto/solicitations/WNaN/, March 2006.

[3] P. Marshall, “Darpa progress towards affordable, dense, and content focused tactical edge networks,” in IEEE Military Communications Conference, MILCOM 2008, November 2008, pp. 1–7.

[4] C. Adjih, T. Clausen, P. Jacquet, A. Laouiti, P. Minet, P. Muhlethaler, A. Qayyum, and L. Viennot, “Optimized link state routing protocol (olsr),” IETF RFC 3626, October 2003.

[5] M. Gerla, X. Hong, and G. Pei, “Fisheye state routing protocol (fsr) for ad hoc networks,” IETF Internet-Draft (expired), draft-ietf-manet-fsr-03, June 2002.

[6] C. Adjih, E. Baccelli, T. Clausen, P. Jacquet, and G. Rodolakis, “Fish eye olsr scaling properties,” Journal of Communication and Networks, vol. 6, no. 4, pp. 343–351, December 2004.

[7] S. Chinara and S. K. Rath, “A survey on one-hop clustering algorithms in mobile ad hoc networks,” Journal of Network and Systems Management, vol. 17, no. 1-2, pp. 183–207, 2009.

[8] L. Villasenor-Gonzalez, Y. Ge, and L. Lamont, “Holsr: a hierarchical proactive routing mechanism for mobile ad hoc networks,” Communications Magazine, IEEE, vol. 43, no. 7, pp. 118–125, July 2005.

[9] A. P. Chandrakasan, A. C. Smith, W. B. Heinzelman, and W. B. Heinzelman, “An application-specific protocol architecture for wireless microsensor networks,” IEEE Transactions on Wireless Communications, vol. 1, no. 4, pp. 660–670, 2002.

[10] H. Chen and S. Megerian, “Cluster sizing and head selection for efficient data aggregation and routing in sensor networks,” in IEEE Wireless Communications and Networking Conference, WCNC 2006, vol. 4, April 2006, pp. 2318–2323.

[11] P. Koshy and S. V. Raghavan, “An information theoretic approach to cluster head selection in mobile ad hoc networks,” in IFIP-TC6 International Conference on Mobile and Wireless Communications Network, MWCN 2003, October 2003, pp. 120–123.

[12] D. Xia and N. Vlajic, “Near-optimal node clustering in wireless sensor networks for environment monitoring,” in IEEE 21st International Conference on Advanced Networking and Applications, AINA’07, May 2007, pp. 632–641.

[13] C. Konstantopoulos, D. Gavalas, and G. Pantziou, “Clustering in mobile ad hoc networks through neighborhood stability-based mobility prediction,” Computer Networks: The International Journal of Computer and Telecommunications Networking, vol. 52, no. 9, pp. 1797–1824, 2008.

[14] T.W. Haynes, S. Hedetniemi, and P. Slater, “Fundamentals of Domination in Graphs (Pure and Applied Mathematics (Marcel Dekker)).” CRC, January 1998.

[15]A. Amis, R. Prakash, T. Vuong, and D. Huynh, "Max-min d-cluster formation in wireless ad hoc networks," in 19th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 2000, vol. 1, March 2000, pp. 32–41.
[16]C.-S. Nam, Y.-K. Ku, J.-W. Yoon, and D.-R. Shin, "Cluster head selection for equal cluster size in wireless sensor networks," in International Conference on New Trends in Information and Service Science, NISS'09, June 2009, pp. 618–623.
[17]M. Chatterjee, S.-K. Das, and D. Turgut, "WCA: a weighted clustering algorithm for mobile ad hoc networks," Journal of Cluster Computing, vol. 5, pp. 193–204, 2002.
[18]F. Kuhn, T. Moscibroda, and R. Wattenhofer, "Fault-tolerant clustering in ad hoc and sensor networks," in IEEE 26th International Conference on Distributed Computing Systems, ICDCS'06, July 2006, pp. 68–77.
[19]U. Feige, "A threshold of ln n for approximating set cover," Journal of the ACM, vol. 45, pp. 314–318, 1998.
[20]R. M. Karp, "Reducibility among combinatorial problems," in Complexity of Computer Computations, R. E. Miller and J. W. Thatcher, Eds. Plenum Press, 1972, pp. 85–103.
[21]T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, "Introduction to Algorithms," 2nd ed. The MIT Press, September 2001.
[22]V. Chvátal, "A greedy heuristic for the set-covering problem," Mathematics of Operations Research, vol. 4, no. 3, pp. 233–235, 1979.
[23]L. Jia, R. Rajaraman, and T. Suel, "An efficient distributed algorithm for constructing small dominating sets," Journal of Distributed Computing, vol. 15, no. 4, pp. 193–205, 2002.
[24]T. H. Clausen, C. Dearlove, and J. W. Dean, "Mobile ad hoc network (manet) neighborhood discovery protocol (nhdp)," IETF Internet-Draft (work in progress), draft-ietf-manet-nhdp-12, March 2010.
[25]J. Edmonds, "Paths, trees, and flowers," Canadian Journal of Mathematics, vol. 17, pp. 449–467, 1965.

Dang Nguyen Dr. Nguyen is a research scientist at the Communications Research Centre, Ottawa, Canada. He has obtained his Master degree in 2003 and his PhD in 2006, both in Computer Science at University of Paris VI (Pierre and Marie Curie), France. He has undergone various research projects at INRIA (France), Orange's Labs (France) and CRC (Canada) under major government grants and contracts. His research interest includes quality of service in MANETs, trust-based security and cryptography.

Pascale Minet Dr. Minet is a senior researcher at INRIA Rocquencourt, France. She is the deputy-head of the HIPERCOM team-project and co-author of the OLSR protocol (IETF's RFC 3626). Her research interest includes quality of service and multicast in MANETs, wireless sensor networks with energy efficiency, node activity scheduling and energy-aware routing.

Thomas Kunz Dr. Kunz received a double honors degree in Computer Science and Business Administration in 1990 and the Dr. Ing. degree in Computer Science in 1994, both from the Darmstadt University of Technology, Federal Republic of Germany. He is currently a Professor in Systems and Computer Engineering at Carleton University. His research interests are primarily in the area of wireless and mobile computing. The main thrust is to facilitate the development of innovative next-generation mobile applications on resource-constraint, hand-held devices, exploring the required network architectures (MANETs, wireless mesh networks, wireless sensor networks), network protocols (routing, Mobile IP, QoS support), and middleware layers. He authored or co-authored close to 150 technical papers, received a number of awards, and is involved in national and international conferences and workshops.

Louise Lamont is the Research Manager for the Mobile ad hoc and Sensor Networking Group at the Communications Research centre. In this position Louise is responsible for identifying novel research areas for study at CRC and for proposing, implementing and securing funds for new projects in support of major client requirements such as DND. She manages several state-of-the-art laboratories for the conduct of research as well as technical demonstration in the area of mobile ad hoc and sensor networks. She is also responsible for establishing and maintaining liaison, collaboration and partnership with R&D groups at CRC and with external national and international organizations.