Computing connected dominated sets with multipoint relays

Cédric Adjih, Philippe Jacquet, Laurent Viennot

Computing connected dominated sets with multipoint relays

Cedric Adjih[*] Philippe Jacquet[*] Laurent Viennot[*]

August 3, 2004

Abstract

Multipoint relays offer an optimized way of flooding packets in a radio network. However, this technique requires the last hop knowledge: to decide whether or not a flooding packet is retransmitted, a node needs to know from which node the packet was received. When considering broadcasting at IP level, this information may be difficult to obtain. We thus propose a scheme for computing an optimized connected dominating set from multipoint relays. Proof of correctness and simulations are given for all these broadcasting mechanisms.

Key-words: multipoint relays, connected dominating set, ad hoc network

1 Introduction

Ad hoc networks [3] offer new paradigms for routing. Unicast routing has been widely studied. On the other hand, broadcasting in an ad hoc network is still an open issue. The main focus here will be on optimizing the bandwidth utilization for broadcasting a packet. Wireless interfaces have an inherent capacity for broadcasting, *i.e.* with one emission, a node can reach all the nearby nodes. However, it is hard in practice to benefit from this capacity in order to avoid redundant retransmissions.

Another problem is to avoid loops when broadcasting. We will assume in the following that some mechanism allows a node to know if it has already received a packet or not. This can usually be achieved by using sequence numbers or computing signatures of packets.

The simplest way of broadcasting a packet to all nodes in an ad hoc network is basic flooding, that we call pure flooding.

> **Pure flooding forwarding rule:** a node retransmits if it has not already received the packet.

[*]INRIA Rocquencourt, Domaine de Voluceau, 78153 Le Chesnay, France, {`Cedric.Adjih, Philippe.Jacquet, Laurent.Viennot`}`@inria.fr`

Clearly all nodes reachable from the source will receive the packet. However, every node will retransmit the packet when it is possible, as we will see, to greatly reduce the number of retransmissions.

Several techniques have been proposed to optimize flooding with two hop neighborhood knowledge (*i.e.* a node knows its neighbors and the neighbors of its neighbors). Link state routing protocols usually provide this information. It is usually obtained by regularly emitting so-called Hello packets containing lists of neighbors.

From graph theory point of view, the set of nodes that will retransmit a given broadcast packet (including the source) must form a connected dominating set. A dominating set is a set of nodes such that any node in the network is neighbor of some element of the set. It is connected if the subgraph formed by this set is connected. The connectedness of the dominating set insures that all nodes of the connected dominating set will receive the packet and will thus be able to retransmit it. The dominating set property insures that all nodes will receive the packet (assuming no transmission error).

Computing a minimum size connected dominating set is NP hard. Moreover, the selection of the connected dominating set must be distributed. Based on two hop neighborhood knowledge, a node has to decide whether or not it is in the dominating set. Some heuristics based on neighborhood inclusions have been proposed [14, 4, 9, 5, 11, 12]. If a connected dominating set has been elected, then the forwarding rule becomes:

> **Connected dominating set forwarding rule:** a node retransmits if it has not already received the packet and it is in the connected dominating set.

On the other hand, the multipoint relay technique has been proposed to optimize flooding when the last hop information is given [6, 2, 7]. The idea behind this technique is to compute some kind of local dominating sets. Each node computes a multipoint relay set with the following properties:

- the multipoint relay set is included in the neighborhood of the node, the element of the multipoint relay set are called *multipoint relays* (or MPR for short) of the node,
- each two hop neighbor of the node has a neighbor in the multipoint relay set (we say that some multipoint relay *covers* the two hop neighbor).

The multipoint relay set plus the node forms a dominating set of the two hop neighborhood of the node.

More formally, let $\mathcal{N}(E)$ denote the set of all nodes that are in a given set E or have a neighbor in E. We say that E *covers* a set F when $F \subset \mathcal{N}(E)$. Let $\mathcal{N}_1(E) = \mathcal{N}(E) - E$ denote the nodes at distance 1 from E and $\mathcal{N}_2(E) = \mathcal{N}(\mathcal{N}(E)) - \mathcal{N}(E)$ denote the nodes at distance 2 from E. When $E = \{x\}$ contains only one node, $\mathcal{N}_1(\{x\})$ is the neighborhood of x and $\mathcal{N}_2(\{x\})$ is the two hop neighborhood of x. A MPR set M of a node x is thus any subset of $\mathcal{N}_1(\{x\})$

such that $\mathcal{N}_2(\{x\}) \subset \mathcal{N}(M)$. We say that M is a subset of neighbors that covers the two hop neighborhood of x or equivalently M which is a dominating set of the subgraph induced by $\mathcal{N}(\mathcal{N}(\{x\}))$.

We call multipoint relay selector of a node x a node which has selected node x as multipoint relay.

It can then be proven that the following forwarding rule allows to reach all the nodes in the network (assuming no transmission error):

> **Multipoint relay flooding forwarding rule:** a node retransmits if it has not already received the packet and it is a multipoint relay of the last emitter.

The smallest the multipoint relay sets are, the fewer retransmissions will occur. It is NP hard to compute a multipoint relay set with minimum size [7] but there exists good heuristics based on preferring neighbors with large degree as multipoint relays.

In this paper, we propose an algorithm for computing a connected dominating set based on multipoint relays. The only knowledge assumed for a given node is two hop neighborhood and the list of neighbors that have selected the node as multipoint relay (such neighbors are called multipoint relay selectors). This information can be contained in hello packets that nodes periodically broadcast to their neighbor in order to monitor links validity. The algorithm does not need any distributed knowledge of the global network topology. These assumptions make the algorithm very attractive for mobile ad hoc networks since it needs just local updates at each detected topology change.

This paper first gives some insight about MPR sets computation. We then compare MPR connected dominating set broadcasting to MPR flooding and to other classical connected dominating set heuristics. Finally we include proofs of correctness for all these techniques.

2 Computing MPR Sets

Two interesting ways of computing MPR sets have been proposed [1, 8].

2.1 Greedy MPR set computation

In practice the following greedy algorithm [1] works very well for computing the MPR set of a node x. If v is a two hop neighbor of x, we call connecting neighbors the nodes that are connected to both x and v.

We start with an empty MPR set.

- Step 1: Find the two-hop neighbors that have only one connecting neighbor. Put in the MPR set the connecting neighbors in $\mathcal{N}(x)$ that connect u to these two-hop neighbors. (Notice that any MPR set must contain these connecting neighbors.)

- Step 2: Add in the MPR set the neighbor node that covers the largest number of two-hop neighbors of x that are not yet covered by the current MPR set.
- Repeat Step 2 until all two-hop neighborhood is covered.

It has been shown [7] that the greedy algorithm provides an MPR set whose size is at most $\log m$ larger than the optimal MPR set, where m is the maximum degree of a node. The greedy algorithm computation takes $O(m^2)$ time. (Computing the sets of connecting neighbors of every two-hop neighbor takes $O(m^2)$ time. The number of uncovered two-hop neighbors covered by each neighbor can be updated at each step with overall complexity $O(m^2)$.)

2.2 Min-id MPR set computation

We describe here an algorithm called the *min-id* algorithm that consists into selecting the nodes in the increasing order of their ID's (or any arbitrary increasing order). It is similar to the "parent" computation proposed in [8].

Start with an empty MPR set.

- Check the neighbor nodes in the increasing order of their identifiers. If the current node covers a two-hop neighbor which was not yet covered by the current MPR set, then add the current node to the MPR set.

Although the min-id algorithm is far from optimal (there is no bound with regard to the optimal MPR set size), it also has the advantage that the node can detect by itself whether or not it belongs to the MPR set of a neighbor [8], but this computation takes $O(m^3)$ time. The set of nodes that select a given node x as MPR is called the MPR selector set of x. The above reverse MPR selection algorithm for node u operates as follow:

Start with an empty MPR selector set.

- For each pair of neighbor nodes x and v, compute the set of nodes that neighbors of both x and v by intersecting the neighborhood of x and v. If node u has the smallest identifier among them, then x and v are both added to the MPR selector set.

Both algorithms clearly agree on MPR relations.

2.3 Simulation model

Our first set of simulations will compare the two above algorithms for computing MPR sets. We take as model network of 2,000 nodes randomly and uniformly in a square field $L \times L$. Two of nodes are in range of each other when the distance is smaller than a fixed radius R. This model defines the disk unit graph which is classically used for modeling radio connectivity in an open field with a fixed radio range.

We call m the central node degree which is defined as $m = 2000\frac{\pi R^2}{L^2}$. It is the average neighborhood size of a node in the center of the square field. We

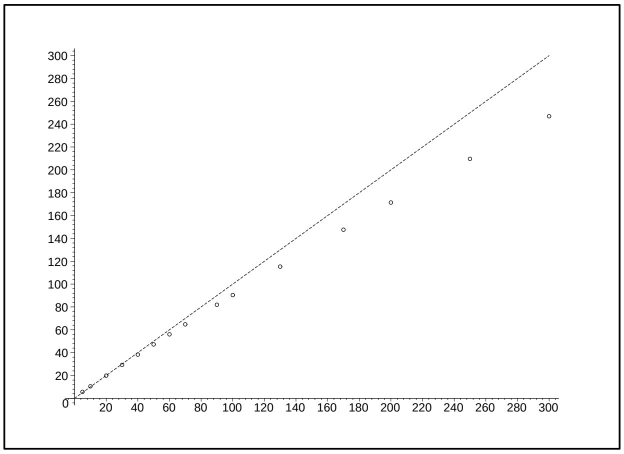

Figure 1: Average neighborhood size versus central node degree.

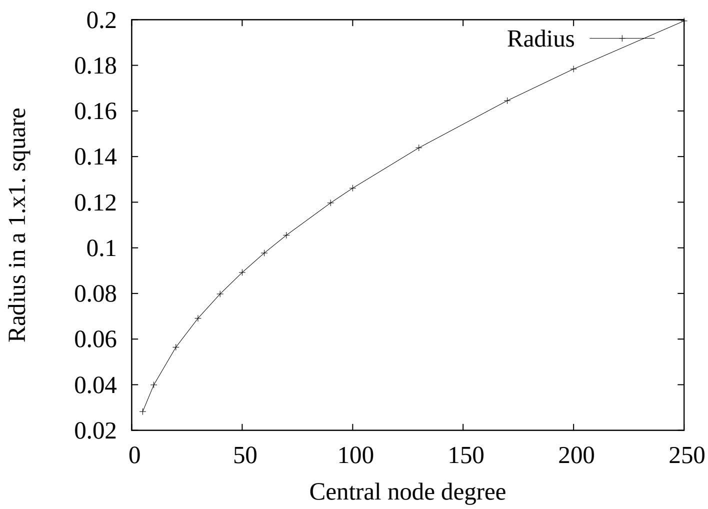

Figure 2: Radio range or radius versus central node degree.

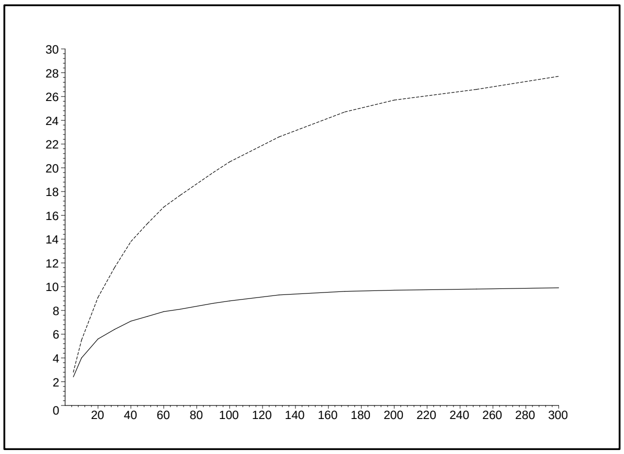

Figure 3: Average MPR size with greedy algorithm (plain) and with min-id algorithm (dashed) versus central node degree.

run different simulations with different values of R in order to let m vary. The average neighborhood size for any node is smaller than m since some nodes have their neighborhood cut by the border of the field. Figure 1 shows the actual average neighbor size with respect to density m. The border effect becomes apparent when $m > 100$.

2.4 Comparing greedy and min-id MPR algorithms

Figure 3 displays the average number of MPR per node using greedy algorithm and min-id algorithm. It turns out that the number of MPR nodes is much smaller than the average neighborhood size, illustrating the performance of the MPR selection. Anyhow the greedy algorithm is more efficientt than the min-id algorithm.

Figure 4 shows the average fraction of nodes in the network that retransmit in an MPR flooding. Notice that the performance ratio between greedy algorithm and min-id algorithm is similar to the ratio between average MPR set sizes. (Some artifact is observed for low densities because of unconnected network situations.)

Figure 5 shows the total number of nodes in the network that are MPR of at least one node in the network. Notice the surprising results that greedy algorithm and min-id algorithm provides very similar results although the average numbers of MPR per node greatly differ. The reason is that the min-id algorithm counterbalances its low performance by forcing nodes to be MPRs

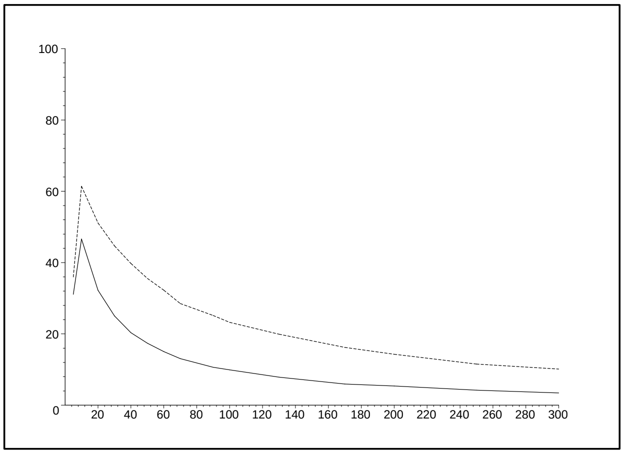

Figure 4: Average fraction of nodes in the network implied in an MPR flooding, with greedy algorithm (plain) and with in-id algorithm (dashed), versus central node degree.

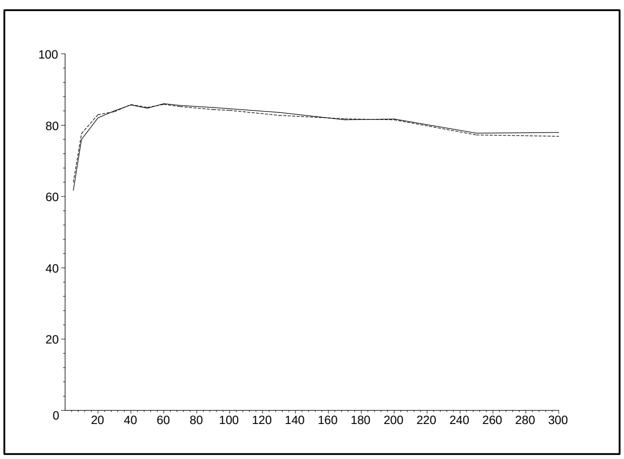

Figure 5: Average fraction of nodes which are MPR of at least one node in the network, with greedy algorithm (plain) and with in-id algorithm (dashed), versus central node degree.

of several nodes. However the dominating set made by all the MPRs is not performant (too many nodes).

3 Computing a Connected Dominating Set

We now propose a connected dominating set election algorithm based on multipoint relays that we call *MPR-CDS*.

3.1 MPR Connected Dominating Set

To allow coordination between nodes, our algorithm requires the knowledge of a total order of the nodes. One can possibly use the smallest IP address of a node as an ID, and we will say that a node is smaller than an other if it has a smaller ID. Any total order on which all nodes agree can be used.

> **MPR-CDS rule:** A node decides that it is in the connected dominating set if and only if:
>
> - the node is smaller than all its neighbors (rule 1),
> - or it is multipoint relay of its smallest neighbor (rule 2).

The correctness of this procedure is shown in Section 4.1. Theoretical relationship with MPR-flooding is also shown. We first show how MPR flooding and MPR-CDS are tightly related on simulation results (with the same model as before).

3.2 Comparing MPR flooding with MPR-CDS

Figure 6 shows the fraction of nodes selected by the MPR-CDS algorithm (rule 1 and rule 2) compared to the fraction of nodes retransmitting in a MPR flooding. To better understand the relationship between MPR flooding and MPR-CDS, a "rule 2" curve illustrates the fraction of nodes selected by applying only rule 2 of the algorithm.

As we can see the MPR flooding curve lies in-between the two curves for MPR-CDS and rule 2 alone. Intuitively, MPR flooding corresponds to applying only rule 2 with respect to the order of emissions as we will see more formally in Section 4.4. However, the order of emission seems slightly less favorable than a random order. (Each re-transmitter waits a random delay before emitting.)

3.3 Generalized Wu and Li CDS algorithm

To evaluate our MPR-CDS algorithm, we compare it to the best CDS heuristic to our knowledge. (Recall that we restrict our framework to the knowledge of two-hop neighborhood.) Many heuristics have been proposed for computing a CDS in an ad hoc network [14, 4, 9, 5, 11, 12]. However, most of them are

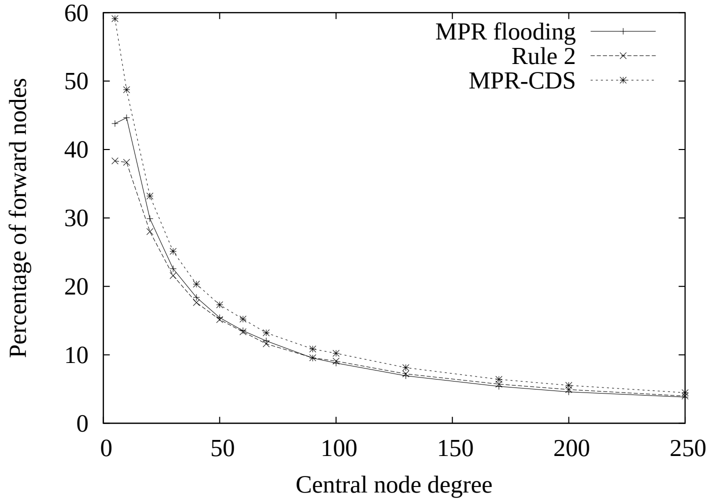

Figure 6: Average fraction of nodes implied in a MPR flooding or a MPR-CDS

inspired from the seminal elimination scheme of Wu and Li [14]. See [13] for an overview of these techniques.

The best results are obtained with the following straightforward generalization of Wu and Li rules:

> **GWL-CDS rule:** A node u considers the set M of neighbors with ID larger than the ID of u. u considers it is in the connected dominating set if and only if one of the two following rules fail:
>
> - M is connected,
> - M covers $\mathcal{N}(u)$ (*i.e.* $\mathcal{N}(u) \subset \mathcal{N}(M)$).

We call the resulting set the *GWL-CDS*. The test whether M is connected can be extended by using two-hop neighbors with greater ID than u. The above algorithm is an undirected version of "rule k" algorithm [12]. It is also very close to [10]. However, all these techniques have similar performances [13].

3.4 Comparing MPR-CDS and GWL-CDS

Figure 7 compares MPR broadcasting techniques (MPR flooding and MPR-CDS) with GWL-CDS. The three techniques behave similarly. However, GWL-CDS behaves slightly better with regard to the number of emissions. Notice that these simulations hold for 2000 nodes.

Figure 8 displays the same simulations for a 100 nodes network. In that case, MPR-CDS and GWL-CDS behave equally well, and MPR flooding is slightly better. This can be explained by an edge effect of the simulation field. With

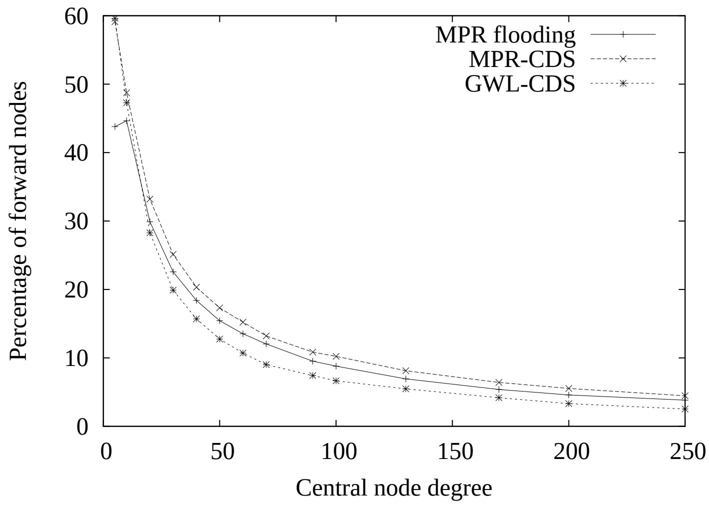

Figure 7: Average fraction of nodes in the network implied in an MPR flooding, in a MPR-CDS and in a GWL-CDS (with a 2000 nodes network).

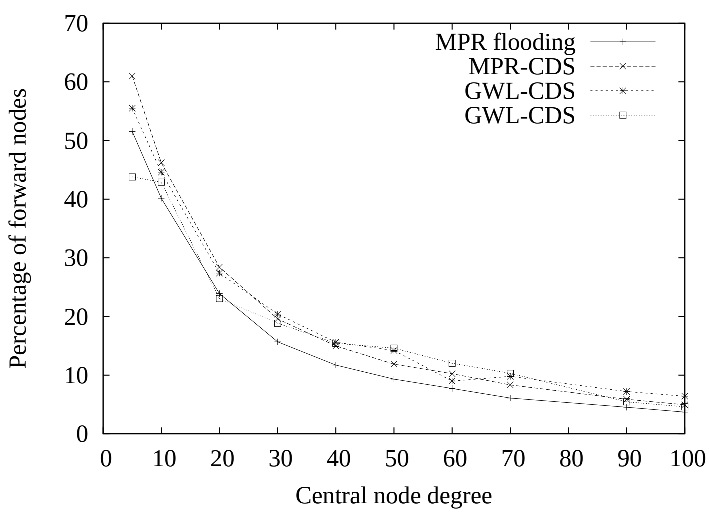

Figure 8: Average fraction of nodes in the network implied in an MPR flooding, in a MPR-CDS and in a GWL-CDS (with a 100 nodes network).

fewer nodes, the same central node degree implies a larger radio range with respect to the field dimensions. MPR selection seems to react very well to such discontinuity in node repartition.

4 Correctness of proposed mechanisms

4.1 Correctness of MPR-CDS

Let us call D the set of all nodes that have decided to be in the connected dominating set. The smallest node of the network is clearly in D by rule 1. Let C be the connected component of the smallest node in the subgraph induced by D. We are going to show that C is a dominating set for the network (assuming that the network is connected of course). This will prove in particular that any node in D has a neighbor in C, implying that C indeed equals D. This will thus prove that D is connected on the one hand and that D is a dominating set on the other hand.

Assume by contradiction that C is not a dominating set. There must exist some nodes that are not in $\mathcal{N}(C)$. Consider the set V of nodes connecting some node in C to some node in $\overline{\mathcal{N}(C)}$, the complementary of $\mathcal{N}(C)$. V is the set of nodes which have at least one neighbor in C and at least one neighbor in $\overline{\mathcal{N}(C)}$. As the network is connected, our assumption implies that V is not empty. Notice that $V \cap C = \emptyset$ by construction. We now consider the smallest node m in $\mathcal{N}(V)$.

- Either m is in $\overline{\mathcal{N}(C)}$. As $m \in \mathcal{N}(V)$, there exists a neighbor v of m in V. Let c be a neighbor of v in C. Consider the multipoint relay set of m: as c is a two hop neighbor of m, there must exist some multipoint relay r of m which is a neighbor of c. Notice that r must be in V. As the smallest neighbor of r is m, r should have elected itself in D by rule 2, contradicting $V \cap C = \emptyset$.
- Either m is in $\mathcal{N}(C)$ which can be partitioned in V and $\mathcal{N}(C) - V$:
  - If m is in V, m should have elected itself as being in D since all its neighbors are greater. This is again a contradiction with $V \cap C = \emptyset$.
  - On the other hand if m is not in V, it cannot have any neighbor in $\overline{\mathcal{N}(C)}$. Let v be a neighbor of m in V, and let x be a neighbor of v in $\overline{\mathcal{N}(C)}$. As m has no neighbor in $\overline{\mathcal{N}(C)}$, x is not a neighbor of m and some multipoint relay r of m must be connected to x. As m cannot have any neighbor in $\overline{\mathcal{N}(C)}$, r must be in $\mathcal{N}(C)$. As r has a neighbor in $\overline{\mathcal{N}(C)}$, it is in V. The smallest neighbor of r is thus m implying that r should be in D by rule 2, contradicting again $V \cap C = \emptyset$.

In all cases we get a contradiction. C thus have to be a dominating set. As mentioned before, this implies that $D = C$ is a connected dominating set.

4.2 Correctness of GWL-CDS

For the sake of clarity, we include a proof of correctness for GWL-CDS algorithm.

Let $u_1, \ldots, u_n$ denote the nodes of a connected network sorted by increasing ID. Let D be the set of nodes that elect themselves as being part of the GWL-CDS with the above rules. We set $D_i = D \cup \{u_{i+1}, \ldots, u_n\}$. We show by induction on i that D_i is a connected dominating set.

D_0 contains all nodes and is clearly connected and dominating.

Suppose D_{i-1} is connected and dominating. If u_i does not eliminate itself, $D_i = D_{i-1}$ is also connected and dominating. Otherwise, the neighbors $v_1, \ldots, v_k$ of u_i with greater ID are connected and cover $N(u_i)$. As $v_1, \ldots, v_k$ have a larger ID than u_i they are all in $\{u_{i+1}, \ldots, u_n\}$ and thus in D_i. $D_i = D_{i-1} - \{u_i\}$ is still dominating since any neighbor of u_i is dominated by some node among $v_1, \ldots, v_k$. D_i is still connected since any pair of node connected via u_i can still be connected by a path in $v_1, \ldots, v_k$.

4.3 MPR flooding correctness

When a multipoint relay flooding is carried out, the set of nodes that retransmit the packet (including the source) forms a connected dominating set. We recall the proof of that fact for the paper to be self contained.

The nodes that retransmit are clearly connected. It is thus sufficient to show that any node receives the message. Suppose by contradiction that some node y does not receives the message. Consider the closest nodes to y (in number of hops) that receive the message. Let C denote this set and let d be the distance in number of hops between y and nodes in C. (C is not empty as soon as we assume that some path in the network connects the source of the MPR flooding to y.) Let B be the set of re-transmitters in $\mathcal{N}(C)$. All these nodes must be at distance $d+1$ from y by our contradiction hypothesis. Let m be the first emitter in B. m has a neighbor in C and a two-hop neighbor at distance $d-1$ from y. m must thus have elected an MPR r to cover this two-hop neighbor. r must be in C and should have retransmitted according to the MPR forwarding rule. This contradiction terminates the proof.

4.4 Comparison between MPR flooding and MPR-CDS

We are going to show that the connected dominating set obtained by a multipoint relay flooding can be obtained by our algorithm if the right order on the nodes is taken *a posteriori*.

Consider one multipoint relay from a source s and let R denote the set of re-transmitters including s. We define a total order on nodes as follows:

- s is smaller than any other node,
- any node in R is smaller than any node not in R,

- for $x, y \in R$, $x < y$ if and only if x has emitted before y,
- the nodes not in R are ordered arbitrarily.

Let us show that our algorithm computes R as connected dominating set. Let D denote the set computed by our algorithm. We have to show $D = R$. First notice that $s \in D$ by rule 1. Consider any node $x \in R - \{s\}$. Since x has emitted, the first time it received the packet from some neighbor y, he was multipoint relay from y and decided to retransmit. y is clearly the smallest neighbor of x, implying $x \in D$ by rule 2. This proves $R \subset D$.

Now consider some node $x \in D$, and let m be the smallest node in $\mathcal{N}(\{x\})$. As proven before, R is a dominating set. m is thus in R. Either $m = x$ proving $x \in R$ or $m \neq x$, implying that x was elected by rule 2: x is a multipoint relay of m. As m was the first neighbor of x re-transmitting the packet, x must have decided to retransmit, and again $x \in R$. This proves $D \subset R$, achieving the proof of $D = R$.

The multipoint relay flooding forwarding rule can be seen as a particular case of our connected dominating set algorithm where the only node elected by rule 1 is the source. This is thus not surprising that multipoint relay flooding gives better results when nodes are ordered arbitrarily since a node is elected by rule 1 with probability $1/(\delta)$ (where δ is its degree). The average number of nodes elected by rule 1, is then approximately N/Δ when all nodes have a degree close to the average degree Δ.

5 Conclusion

We have proposed a connected dominating set election algorithm based on multipoint relays. This algorithm shows good results compared to other CDS heuristics. An interesting feature of multipoint relay flooding is to allow “willingness” and “coverage” tuning [1]. Willingness allow nodes to declare if they want or not to relay broadcast packets. The coverage allows to find a good compromise between optimization and reliability by tuning a degree of redundancy in MPR flooding. Important future work resides in allowing such facilities in connected dominating sets algorithms.

References

[1] T. Clausen, P. Jacquet (editors), C. Adjih, A. Laouit i, P. Minet, P. Muhlethaler, A. Qayyum, and L.Viennot. Optimized link state routing protocol (olsr). RFC 3626, October 2003. Network Working Group.

[2] ETSI STC-RES10 Committee. Radio equipment and systems: High performance radio local area network (hiperlan) type 1, functional specifications, June 1996. ETS 300-652.

[3] M. S. Corson and J. Macker. Mobile ad hoc networking (MANET): Routing protocol performance issues and evaluation considerations. RFC 2501, January 1999.

[4] S. Datta, I. Stojmenovic, and J. Wu. Internal node and shortcut based routing with guaranteed delivery in wireless networks. In *Proc. IEEE Int. Conf. on Distributed Computing and Systems*, April 16-19 2001. Phoenix, AR.

[5] Sudipto Guha and Samir Khuller. Approximation algorithms for connected dominating sets. In *European Symposium on Algorithms*, pages 179–193, 1996.

[6] Philippe Jacquet, Pascale Minet, Paul Muhlethaler, and Nicolas Rivierre. Increasing reliability in cable-free radio LANs: Low level forwarding in HIPERLAN. *Wireless Personal Communications*, 4(1):65–80, January 1997.

[7] A. Laouiti, A. Qayyum, and L. Viennot. Multipoint relaying: An efficient technique for flooding in mobile wireless networks. In *35th Annual Hawaii International Conference on System Sciences (HICSS'2001)*. IEEE Computer Society, 2001.

[8] R.G. Ogier, F.L. Templin, and M.G. Lewis. Topology dissemination based on reverse-path forwarding (tbrpf). RFC 3684, February 2004. Network Working Group.

[9] I. Stojmenovic, M. Seddigh, and J. Zunic. Internal node based broadcasting algorithms in wireless networks. In *Proc. IEEE Hawaii Int. Conf. on System Sciences*, January 2001.

[10] J. Sucec and I. Marsic. An efficient distributed network-wide broadcast algorithm for mobile ad hoc networks. CAIP Technical Report 248, Rutgers University, september 2000.

[11] P.-J. Wan, K. M. Alzoubi, and O. Frieder. Connected dominating set in wireless ad hoc networks. In *IEEE INFOCOM*, June 2002. New York.

[12] J. Wu and F. Dai. Distributed dominant pruning in ad hoc wireless networks. In *Proceedings of IEEE International conference on Communications (ICC)*, may 2003.

[13] J. Wu and F. Dai. A generic distributed broadcast scheme in ad hoc wireless networks. In *ICDCS*, 2003.

[14] J. Wu and H. Li. On calculating connected dominating set for efficient routing in ad hoc wireless networks. In *Proc. DIAL M*, pages 7–14, August 1999. Seatle, USA.