Efficient Broadcasting with Guaranteed Coverage in Mobile Ad Hoc Networks *

Jie Wu and Fei Dai
Department of Computer Science and Engineering
Florida Atlantic University
Boca Raton, FL 33431

Abstract

We study an efficient broadcast scheme in mobile ad hoc networks (MANETs). The objective is to determine a small set of forward nodes to ensure full coverage. We first study several methods that guarantee coverage when the *local view* of each node on its neighborhood information is updated in a timely manner. Then we consider a general case, where nodes move even during the broadcast process, making it impractical to maintain up-to-date and consistent local views. A formal framework is used to model inaccurate local views in MANETs, where full coverage is guaranteed if three sufficient conditions, connectivity, link availability, and consistency, are met. Three solutions are proposed to satisfy those conditions. First we give a minimal transmission range that maintains the connectivity of the virtual network constructed from local views. Then we use two transmission ranges, one for neighborhood information collection and the other for actual data transmission, to form a *buffer zone* that guarantees the availability of logical links in the physical network. Finally we propose a mechanism called *aggregated local view* to ensure consistent local views. By these, we extend Wu and Dai's coverage condition for broadcasting in a network with mobile nodes. The effectiveness of the proposed scheme is confirmed via both performance analysis and simulation study.

Keywords: Broadcasting, localized algorithms, mobile ad hoc networks (MANETs), mobility, simulation.

1 Introduction

Broadcasting a packet to the entire network is a basic operation and has extensive applications in mobile ad hoc networks (MANETs). For example, broadcasting is used in the route discovery process in several routing protocols such as DSR [14], AODV [20], ZRP [13], and LAR [16], when advising an error message to erase invalid routes from the routing table, or as an efficient mechanism for reliable multicast in a fast-moving MANET. In MANETs with the promiscuous receiving mode, the traditional blind flooding incurs significant redundancy, collision, and contention, which is known as the broadcast storm problem [29]. Efficient broadcasting in a MANET focuses on selecting a small forward node set while ensuring broadcast coverage.

Broadcast protocols can be classified into *deterministic* and *probabilistic* approaches. The probabilistic approach [29, 12] usually offers a simple solution in which each node, upon receiving a broadcast packet, forwards the broadcast message with probability p. However, the probabilistic approach cannot guarantee full coverage. The deterministic approach guarantees full coverage and can be further classified based on the type of neighborhood information used: *location-information-based* and *neighbor-set-based*. In location-information-based broadcast protocols, location information of neighbors is available, whereas in neighbor-set-based broadcast protocols, only neighbor set information is available. Location information facilitates efficient broadcasting in terms of generating a small forward node set; however, it comes with a cost – location information requires additional hardware such as GPS. Other types of information can also be used which fall in between the above two models: directional information where messages arrive from a certain angle-of-arrival (AOA) and distance information based on the signal strength received. All these models assume some sort of special hardware. In addition, location/direction/distance information may not be accurate. In this paper, we limit our consideration to deterministic broadcast protocols that use neighbor set information only, which corresponds to the weakest assumption on neighborhood information used.

In a broadcast process, each node decides its forwarding status based on given neighborhood information (such information is constructed from the neighbor set of each node), and the corresponding broadcast protocol is called *self-pruning*. In Figure 1, black (white) nodes are forward (non-forward) nodes. Each circle corresponds to a one-hop neighborhood. Any source node is a black node by default. Basically, forward nodes form a connected dominating set (CDS), where each node in the system is either in the set or the neighbor of a node in the set. That is, each white node has at least one black neighbor. However, most existing broadcast schemes assume either
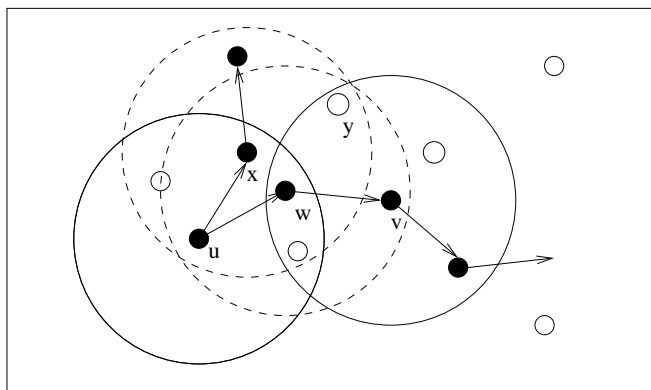
Figure 1: Forward node set in a MANET.

the underlying network topology is static or semi-static during the broadcast process such that the neighborhood information can be updated in a timely manner. The results in [7] show that existing static network broadcast schemes perform poorly in terms of delivery ratio when nodes are mobile. There are two sources that cause the failure of message delivery:

Collision: The message intended for a destination collides with another message. In Figure 1, if messages from nodes w and x collide at node y, node y does not receive any message.

Mobile nodes: A former neighbor moves out of the transmission range of the current node (i.e., it is no longer a neighbor). In Figure 1, when node w moves out of the transmission range of u, the nodes along the branch rooted at w of the broadcast tree will miss the message[1].

Results in [7] show that the effect of collision can be relieved by a very short ($1ms$) forward jitter delay, where a very high ($\geq 99\%$) delivery ratio is achieved in static networks. The majority of delivery failures are caused by mobile nodes. Therefore, in this paper, we focus on delivery failure caused by mobility only. The major challenges in designing a localized broadcast protocol while ensuring broadcast coverage are as follows:

1. The network topology changes over time, even during the broadcast process.

2. The local (1-hop) information is constructed based on "Hello" intervals. To avoid serious collision among "Hello" messages, nodes start their intervals asynchronously, making it difficult to ensure consistent local/global views among nodes.

---

[1]Nodes in the branch may still receive the message, if some adjacent nodes of the branch forward the message.

3. The collection process for k-hop information incurs delay which may not reflect the current network topology when there are mobile nodes, even for small k in localized solutions.

As a consequence, the virtual network constructed from local views of nodes may not be connected (*connectivity issue*), its links may not exist in the physical network (*link availability issue*), and the global view constructed from the collection of local views may not be consistent (*consistency issue*).

In this paper, we first give a sufficient condition for connectivity at the physical network to ensure the connectivity at the virtual network. We then propose a solution using two transmission ranges to address the link availability issue. The neighborhood information as well as the forward node set are determined based on a short transmission range whereas the broadcast process is done on a long transmission range. The difference between these two ranges is based on the update frequency and the speed of node movement. The difference is also used as a new *controllable parameter* to balance broadcast redundancy and broadcast delivery ratio. Although many deterministic broadcast protocols have been proposed with different broadcast redundancies (and collated broadcast delivery ratios), each broadcast protocol has only its "fixed" broadcast redundancy (and broadcast delivery ratio). It is in general hard to control redundancy and delivery for a given broadcast protocol. Note that the forwarding probability in probabilistic broadcasting [30] is also a controllable parameter. However, it is difficult to establish a direct connection between parameter selection and node mobility. Finally, we propose a new mechanism called *aggregated local view* to ensure consistency of the global view. The conservative approach aggregates past k local views in a special way to eliminate potential view inconsistency caused by asynchronous "Hello" intervals and collection process delay for k-hop information at each node.

By providing solutions to the above three issues, we also extend Wu and Dai's coverage condition [31] for broadcasting in a network with mobile nodes. This coverage condition is a sufficient condition for a node to determine its non-forward status based on k-hop neighborhood information (for small k, say 2 or 3) only. However, the coverage condition is only suitable when the topology is static during the broadcast process and neighborhood information is consistent with the current state. Simulation results in this paper show that the proposed scheme improves the coverage significantly.

The main contributions of this paper are as follows:

1. Propose the first localized broadcast protocol that can handle node mobility while ensuring

broadcast coverage.

2. Systematically address the issue of inconsistent local view caused by neighborhood information delay, asynchronous "Hello" intervals, and node mobility.

3. Introduce a new controllable parameter to balance broadcast efficiency and broadcast delivery ratio.

4. Conduct a simulation study to verify the effectiveness of the new approach.

The remainder of the paper is organized as follows: Section 2 provides some preliminaries, especially Wu and Dai's coverage condition, and motivations. Section 3 proposes the mobility control method based on two transmission ranges, presents the consistent global view construction through aggregation of local views, and gives some analytical study and optimization techniques. Simulation results are presented in Section 4. The paper concludes in Section 5.

2   Preliminaries and Motivations

This section starts with some related work on mobility management and, in particular, neighbor set management in a mobile environment. Then an overview of broadcast protocols in MANETs based on self-pruning is given. The focus is on Wu and Dai's coverage condition and six existing protocols as its special cases. Finally, we focus on the limitation of Wu and Dai's coverage condition in dynamic networks to motivate this study.

2.1   Mobility management

The capacity of MANETs is constrained by the mutual interference of concurrent transmissions between nodes. The mobility of nodes adds another dimension of complexity in the mutual interference. Several studies [10, 11] focused the effect of mobility on the network capacity. Camp et al [4] gave a comprehensive survey on mobility models for MANETs. Three popular mobility models include (1) *random walk*, which is a simple mobility model based on random directions and speeds, (2) *random waypoint*, which includes pause time between changes in destination and speed, and (3) *random direction mobility*, which forces hosts to travel to the edge of the simulation area before changing direction and speed. In [23], a *velocity-bounded model* (for pedestrians with

5

mobile nodes in a relatively small area), and an *acceleration-bounded model* (for vehicles of high speed) are given. Other mobility models are discussed in [23], and their impact on performance of routing protocols is discussed in [2].

Very little work has been done in maintaining accurate neighborhood information in a mobile environment without increasing the frequency of "Hello" messages. One exception is [15], where a *stable zone* and a *caution zone* of each node have been defined based on a node's position, speed, and direction information obtained from GPS. Specifically, stable zone is the area in which a mobile node can maintain a relatively stable link with its neighbor nodes since they are located close to each other. Caution zone is the area in which a node can maintain an unstable link with its neighbor nodes since they are relatively far from each other. The drawback of this approach is that it is GPS-based, which comes with a cost. In addition, there is no rigorous analysis on the impact of mobility on the selection of these two zones.

Several papers [3] address the time period that two nodes will remain close enough in proximity for a link between them to remain active. Several routing protocols, associativity-based routing (ABR) [28] and signal stability-based adaptive routing (SSA) [8], have been proposed that select *stable links* to construct a route. In [26], GPS information is used to estimate the expiration time of the link between two adjacent hosts. Recently, several studies have been done on the effect of mobility on routing path [22]. However, no broadcast protocol uses the notion of stable link to evaluate the stability of neighbor set in order to better decide the forwarding status of each node. Although several probabilistic broadcast protocols [12, 30] have been proposed by trading between efficiency (simple design) and coverage (delivery ratio), it is difficult to establish a direct connection between forwarding probability and node mobility.

2.2   Broadcast protocols based on self-pruning

A MANET is usually modelled as an undirected graph $G = (V, E)$, where V is a set of mobile nodes, and E is a set of wireless links. A link exists between two nodes u and v if and only if their physical distance is less than a transmission range r. Wu and Dai [31] proposed a generic efficient broadcast protocol based on self-pruning. In a self-pruning protocol, each node determines its forwarding status based on its local *k-hop information*. Every node is a forward node by default, and becomes a non-forward node (*pruned*) when a sufficient coverage condition holds, as will be discussed later. For a node $v \in V$, its exact k-hop neighbor set, $H_k(v)$, is the set of nodes that is
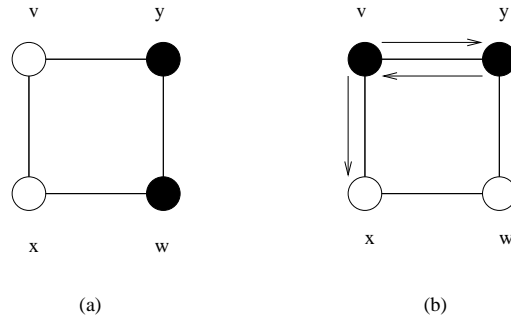
Figure 2: (a) Forward node set without history information (static). (b) Forward node set with upstream history information (dynamic) with node v being the source (visited node).

exactly k-hops away from v, and its k-hop neighbor set, $N_k(v) = \{v\} \cup H_1(v) \cup H_2(v) \cup \ldots \cup H_k(v)$, is the set of nodes that is at most k hops away from v. The k-hop information of v, $G_k(v)$, is the induced graph of $N_k(v)$, excluding links among nodes in $H_k(v)$. For example, links between two nodes exactly 2 hops away are included in 3-hop information, but not in 2-hop information. Each nodes builds its k-hop information by exchanging $(k-1)$-hop information with its neighbors via periodical "Hello" messages. Therefore, k rounds of exchanges of the accumulative neighbor set between neighbors are needed to collect k-hop information at each node.

The "Hello" messages also propagate the *priority* of each node, which could be a static property (e.g., node id) or a dynamic one (e.g., node degree). During a broadcast process, each node may also extract from the incoming broadcast packets a list of *visited nodes* that have forwarded the broadcast packet. Using k-hop information, priority, and visited node information, each node decides its own status, forwarding/non-forwarding, based on the following coverage condition.

Coverage Condition [31]: Node v has a non-forward node status if for any two neighbors u and w, a *replacement path* exists that connects u and w via several intermediate nodes (if any) with either higher priority values than the priority value of v or with visited node status.

Assume node id is used as priority, node x in Figure 2 (a) is a non-forward node based on the coverage condition, because its neighbors, v and w, are connected via a *replacement path* that contains only intermediate nodes (in this case, y) with higher node id than x. Node y is a forward node, because no such replacement path exists. It was proved in [31] that the coverage condition ensures the coverage; that is, the forward nodes, including the source, form a connected dominating set (CDS) of G, if G is connected. Therefore, the broadcast packet is delivered to all nodes in V,
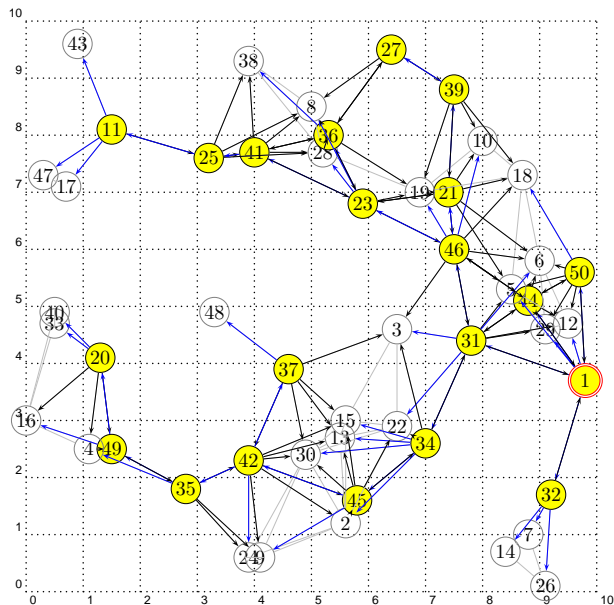
Figure 3: A sample broadcast process. Gray nodes are forward nodes and white nodes are non-forward nodes. Arrows represent receptions of the broadcast packet.

if no packet is lost due to node mobility or MAC layer collision.

A self-pruning protocol is *static* if it does not collect visited node information or use it in the coverage condition; otherwise, it is *dynamic*. In a static protocol, forward nodes are pre-determined before any broadcast process and are source independent. For example, node w in Figure 2 (a) is a forward node in a static protocol, because there is no replacement path to connect neighbors x and y. When node v is the source, there are actually three forward nodes v, w and y. In a dynamic protocol, the forward status is determined during the broadcast process. Each node usually determines its status after a short backoff delay when the node receives the packet for the first time. During the backoff delay, more neighbors may forward the same broadcast packet, and those neighbors become visited nodes. Each broadcast packet carries a list of recently visited nodes, and each node uses the visited node information to enhance the chance of being pruned. The number of forward nodes is usually smaller than in a static protocol, but the selection of forward nodes is usually source dependent. Figure 2 (b) shows the broadcast process in the same network and with the same source node as in Figure 2 (a). This time, node w is a non-forward node, because nodes x and y are connected via node v, which is known by w as a visited node (such information is piggybacked with the broadcast packet). There are only two forward nodes v and y.

In [31], it is assumed that local views of the broadcast specific information (i.e., visited node

8

(a) Broadcast redundancy versus network size   (b) Delivery ratio versus mobility
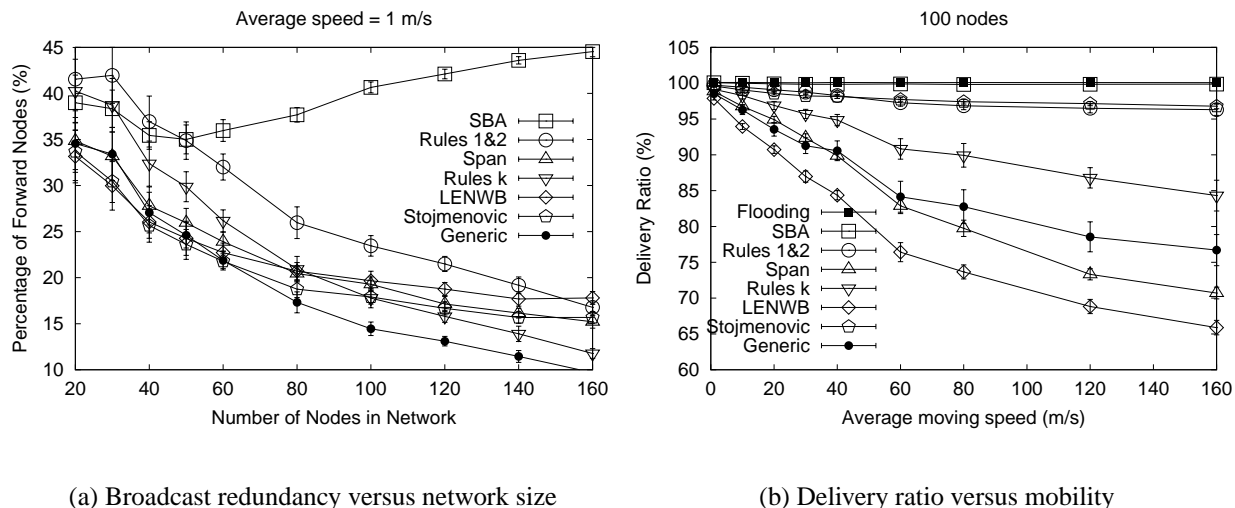
Figure 4: Performance of various broadcast protocols.

information) are dynamic but safe, i.e., an unvisited node will not be mislabelled as visited, and those of the broadcast independent information (i.e., k-hop information and priority) are "static" and accurate during a broadcast process. However, in mobile networks, such "static" information usually changes and causes inaccurate local views. Based on these inaccurate views, full coverage (i.e., 100% delivery ratio) is not guaranteed. In general, the broadcast redundancy and delivery ratio of a self-pruning protocol in a mobile environment is affected by various implementation options, including priority type, "Hello" interval, and backoff delay. Using a static priority, such as node id, is more desirable than a dynamic priority, such as node degree, because a static priority usually achieves higher delivery ratio in mobile networks and has less redundancy in dense networks [7]. Using a small "Hello" interval can provide fresh k-hop information and improve the delivery ratio in a mobile environment, but has the danger of saturating the bandwidth with numerous "Hello" messages. In some self-pruning protocols such as SBA [19], a large backoff delay is essential for achieving high broadcast efficiency. However, a large backoff delay also causes large end-to-end delay. A random jitter delay is also used by each node to avoid collision, but is usually too short to affect the broadcast redundancy or delivery ratio.

Six existing algorithms, including static and dynamic protocols, were shown in [31] to be special cases of the coverage condition. They are Wu and Li's marking process with Rules 1 & 2 (static) [33], Dai and Wu's Rule k (static) [6], Chen et al's Span (static) [5], Sucec and Marsic's LENWB (dynamic) [27], Peng and Lu's SBA (dynamic) [19], and Stojmenovic's algorithm

9

(static/dynamic) [25]. Details of these algorithms are given in the appendix. Figure 3 shows a sample broadcasting based on self-pruning.

As shown in Figure 4, high delivery ratio can be achieved by protocols with high broadcast redundancy, i.e., blind flooding, SBA, and Rules 1&2. A new protocol [31] derived from the coverage condition (labelled as Generic) has the lowest redundancy, but suffers from a low delivery ratio in highly mobile networks. One solution is to use location information as in Stojmenovic's algorithm, which achieves a high delivery ratio with relatively low redundancy. However, using location information incurs extra cost and may not provide accurate prediction on the existence of wireless links. SBA achieves very high delivery ratio in highly mobile networks, but it also has the highest percentage of forward nodes. Note that the percentage of forward nodes in SBA is a function of its backoff delay. A large backoff delay can improve the efficiency of SBA, but can also cause a large end-to-end delay.

3   Proposed Method

This section proposes a mobility control method that addresses connectivity, link availability, and consistency issues. Three sufficient conditions are given: the first one on the connectivity of the physical network to ensure connectivity of the virtual network, the second one on the bound of the range difference to ensure link availability, and the third one on the consistent local views to ensure correct decision made at each node. Two mechanisms, called buffer zone and aggregated local view, are proposed to satisfy the latter two conditions. Finally, we introduce methods to relax these stringent sufficient conditions based on probabilistic analysis and optimization techniques.

3.1   Logical network and broadcast state

In [31], the coverage condition was applied on a static or semi-static physical network. In a semi-static physical network, the physical topology stops to change several "Hello" intervals before a broadcast process, and stays unchanged until the broadcast process completes. For the sake of clarity, we assume node id is used as priority, and define the local view of each node as its k-hop information, which is an approximation of its neighborhood topology.

Two levels of abstraction: The correctness of the coverage condition is based on the assumption that this approximation is an accurate and immediate reflection of the physical topology. In
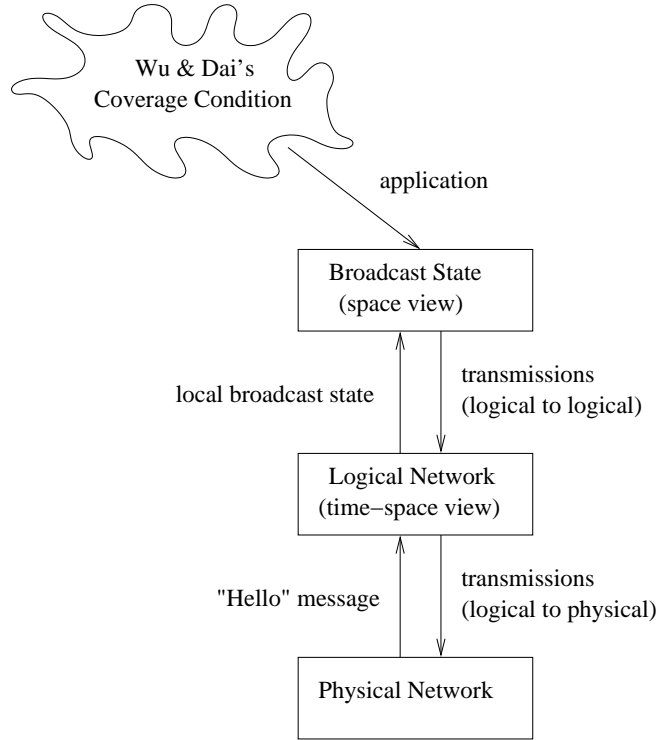
10

Figure 5: The mapping from the logical network and broadcast state to the physical network.

MANETs, however, this assumption can be easily invalidated due to the continuous node mobility. In fact, in order to apply the coverage condition on MANETs with potentially outdated local views, we introduce the concepts of *logical network*, a dynamic virtual network constructed from all local views, and *broadcast state*, a snapshot of local views during a broadcast process, and two levels of abstraction, as shown in Figure 5:

1. Level-1 abstraction: from physical network (time-space view) to logical network (time-space view).

2. Level-2 abstraction: from logical network (time-space view) to broadcast state (space view).

A logical network is the collection of all local views, i.e., a supergraph containing all the nodes and links in local views.

Definition 1 *The* local view, $G'_k(v, t) = (N'_k(v, t), E'_k(v, t))$, *of node* v *is its* k-*hop information collected at time* t. *The* logical network, $G'(t) = (V, E'(t))$, *is the union of all local views at time* t, *where* $E'(t) = \bigcup_{v \in V} E'_k(v, t)$.

11

Both local view and logical network are time sensitive. When the physical topology changes, the change is detected by "Hello" messages and reflected in the logical network. Consider the MANET in Figure 6 (a). We assume that each node has the same "Hello" interval f^2, but each node starts its period asynchronously. Figure 6 (b) shows the update of local views. We label the time each node sends its last "Hello" message before the broadcasting as t_i, and the time for previous "Hello" messages as t_{i-1}, t_{i-2}, and so on. Note that t_i at each node may refer to different physical time. Here each node builds 2-hop information. If node y's "Hello" message is first received by node v between t_{i-2} and t_{i-1} (the "Hello" message propagation is shown in a dotted arrow line), it is added to v's 1-hop neighbor set, which is advertised in v's next "Hello" message at t_{i-1}. That is, link (v, y) is added to local views of nodes v and x. Similarly, link (w, y) is also detected and added to local views of nodes u, w, and x.

Broadcast state is a snapshot of local views of the logical network. For a specific broadcast process, broadcast state forms a virtual static network, upon which the coverage condition is applied.

Definition 2 *A local state $G_k''(v) = (N_k'(v, t_v), E_k'(v, t_v))$ of node v for a broadcast is its local view at the time t_v, when it makes its forwarding/non-forwarding decision. A* global state *$G'' = (V, E'')$, is the union of all local broadcast states, where $E'' = \bigcup_{v \in V} E_k'(v, t_v)$.*

In Figure 6 (b), the time that each individual node makes its decision is marked with a black dot. Note that local broadcast states are taken at the times marked by these black dots, and the global broadcast state is the collection of local broadcast states (marked by the dashed line connecting all black dots). Suppose node v in Figure 6 (a) issues a broadcasting, where nodes y and w become forward nodes. The corresponding broadcast state is shown in Figure 6 (b). The source node v samples its local state at the beginning of the broadcasting, between t_{i-1} and t_i. On receiving the broadcast packet from v, both nodes x and y wait for a random backoff delay and determine their forwarding status. Node x becomes a non-forward node and node y becomes a forward node. Nevertheless, both nodes sample their local state at the time the decision is made. Subsequently, nodes w and u receive the broadcast packet and make their decision based on their local state. The global state is the union of all local states, which are sampled at different physical times during the broadcast process.

---

[2]The condition can also be be relaxed in a controllable way, such as $(1 \pm 0.25)f$ in AODV.

Next we examine the "gap" at each level of abstraction. The gap is caused by various synchronization delays and protocol handshakes at each level of abstraction. We will look potential problems caused by the gap and, in the next subsection, we will present solutions to these problems.

Gap in level 1 abstraction: In order to build k-hop information, each node advertises its $(k-1)$-hop information via "Hello" messages. Each node updates its local view based on received "Hello" messages. Because of asynchronous periodic exchanges among neighboring nodes, the 1-hop neighbor set in a local view at a particular time t does not reflect the actual neighbor set at time t, but the offset is bounded by the "Hello" interval f. In fact, k-hop information is a set that consists of neighborhood information sampled at different times. In general, $H_{i+1}(u)$ was sampled one interval after $H_i(u)$ for $i = 1, 2, ..., k-1$. Clearly, the k-hop information at time t does not reflect the actual neighborhood topology at time t, and the offset is bounded by kf. Suppose the speed of node movement is upper bounded by s. Then sf is the maximum distance a node can move around during a "Hello" interval. The *maximum relative distance* between two nodes in such an interval is $\Delta = 2sf$.

Gap in level 2 abstraction: In a broadcast process based on self-pruning, each node follows three steps: (a) first receipt of broadcast message, (b) backoff delay, and (c) forward/non-forward status decision and transmission (if needed). A *broadcast period* starts from the source sending out the message and ends with the last node deciding its forwarding status. Like [31], it is assumed that the broadcast message propagates quickly and its delay can be ignored. Backoff at intermediate nodes is allowed, but *accumulative backoff* along each path of the broadcast tree is bounded by b, called *broadcast delay*, for each broadcast. Note that b may also include broadcast message propagation delay if such delay cannot be neglected.

3.2 Mobility-sensitive broadcasting

Wu and Dai's coverage conditions can be applied to the global broadcast state and ensures coverage, given that the following three conditions are met: connectivity, link availability, and consistency. The first two conditions resolve the gaps at level 1 and level 2 abstractions, respectively. The third condition ensures the consistent use of local views.

Connectivity: The virtual network that corresponds to the global broadcast state should be connected in order to apply Wu and Dai's condition. The following theorem shows the density re-
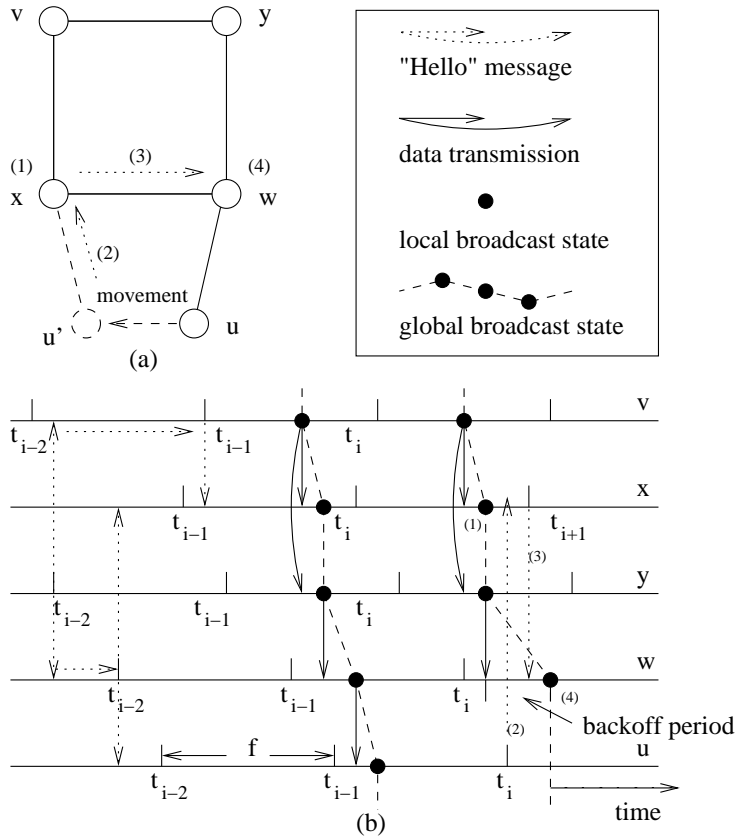
Figure 6: A physical network (a) and the corresponding time-space view (b) during two broadcast processes. The first broadcast process starts after time t_{i-1} and successes. The second one starts after time t_i and fails due to inconsistent local views.

quirement at the physical network for ensuring a connected virtual network.

Theorem 1 *If the physical network with transmission range $r_1 - \Delta'$ is connected under all time, where $\Delta' = 2s(f + b)$, then every virtual network induced from a global broadcast state is connected.*

Proof: Assume the global broadcast state is taken in a broadcast process started at time t. Since the maximum broadcast delay is b, all local states are taken within time period $[t, t + b]$. If the distance of two nodes u and v, $d(u, v) \leq r_1 - 2s(f + b)$ at time $t - f$, then $d(u, v) \leq r_1$ during $[t - f, t + b]$. Suppose u takes its local broadcast state at $t_u \in [t, t + b]$, it must have received v's last "Hello" message in $[t - f, t_u]$. Therefore, link (u, v) exists in u's local broadcast state. Since the global broadcast state consists of all links from local broadcast state, and the network is connected at time $t - f$ in the range of $r_1 - \Delta'$, the corresponding virtual network induced from the global broadcast state is also connected. $\qquad \square$

Theorem 1 poses a rather strict connectivity requirement on the physical network. That is, if the physical network cannot meet the connectivity requirement, the virtual network is not guaranteed to be connected and Wu and Dai's approach will fail. We will discuss later an approach that relaxes the connectivity requirement under the cost of pruning efficiency.

Link availability: Any link in the global broadcast state should still exist in the physical network during the broadcast period. We propose to use two transmission ranges, r_1 and r_2, with $r_1 < r_2$. r_1 is used to collect neighbor set and k-hop information through "Hello" messages, whereas r_2 is used to perform actual transmission. A node that is within the range of r_1 of node u is called a neighbor of u and the collection of such nodes is the neighbor set of u. The set of nodes that are reachable based on r_2 is called *effective neighbor set*. Figure 7 shows the relationship between these two transmission ranges. In this example, v is in u's neighbor set (also in u's effective neighbor set), whereas w is in u's effective neighbor set (but not in u's neighbor set).

Theorem 2 *To ensure the link availability requirement, r_2 should be set so that $\Delta'' \leq r_2 - r_1$, where $\Delta'' = k\Delta + \Delta'$ and k for k-hop information.*

Proof: (sketch) We need to show that any neighbor under the transmission range r_1 when its state is sampled is still an effective neighbor under the transmission range r_2 when the message is sent
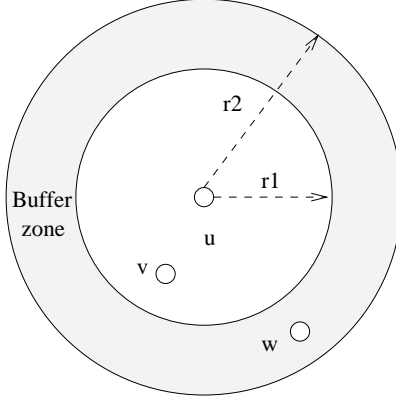
Figure 7: Forward node selection and forwarding process based on two different transmission ranges: r_1 and r_2.

out. The total delay includes k-hop neighbor set collection that takes k intervals, and $(f + b)$ for broadcast and synchronization delay. The former contributes a distance of $k\Delta$ and the latter Δ'. $\square$

The above theorem provides some theoretical foundations for ensuring full coverage. However, the analysis shows only the worst case situation, which rarely occurs. Later we will show that even when $r_2 - r_1$ is much smaller than Δ'', the probability of an undetected link failure is very low. Since most self-pruning protocols have certain degrees of redundancy, it usually takes several link failures to cause a global delivery failure. Therefore, the probability can still be high that a high delivery ratio can be achieved with a relatively small buffer zone width. There is a wide range of potential tradeoffs between broadcast efficiency and broadcast delivery ratio.

The idea of two transmission ranges is to use the "ring", the area bounded by two circles with transmission ranges r_1 and r_2, as a buffer zone to nullify the various bad effects caused by node mobility and transmission delay. However, one bad effect, called inconsistent local views, cannot be nullified no matter how wide the buffer zone is. Inconsistent local views ultimately result in "bad decision" from a node. A decision is *bad* if a node that should forward the message decides on a non-forwarding status.

Consistency: Two local views (or states for a specific broadcasting) of nodes u and v are *inconsistent*, if a link (v, w) exists in u's local view (state), but v does not view w as a 1-hop neighbor. For example, assume the physical topology in Figure 2 changes shortly before the broadcast. The broadcast may fail due to inconsistent views. Figure 8 (a) shows the physical network before the
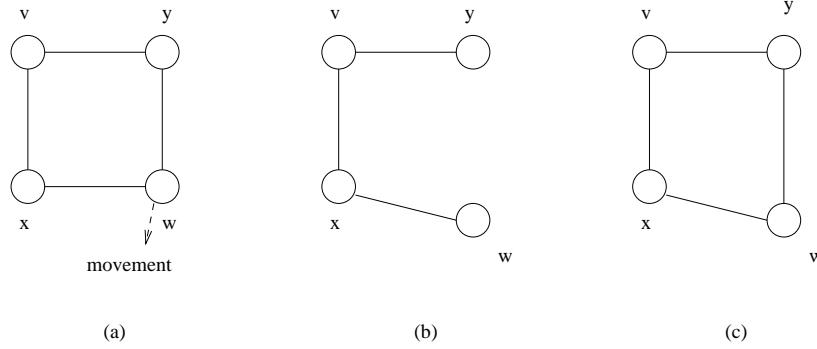
Figure 8: The physical network (a) before and (b) after the movement, and the aggregated local view of node y (c).

change, where node x is a non-forward node because its neighbors v and w are connected via a replacement path (v, y, w). Figure 8 (b) shows the physical network before the broadcast, where y is a non-forward node because w is no longer a neighbor. Node y detects the broken link (y, w) before node x, since y is adjacent to the link whereas x is 2-hop away from the link. Both nodes may take a non-forwarding status in the broadcast, x's decision based on the outdated view and y's based on the updated view. Therefore, node w may never receive the broadcast packet.

We say node v's local view (state) is *consistent* with the logical network (global state) (or simply consistent), if its local view (state) contains all its adjacent links in the logical network (global state). Formally speaking, let $H_1(v)$ denote v's exact 1-hop neighbor set in its local view (state), and $H_1^*(v)$ its exact 1-hop neighbor set in the logical network (global state). The local view (state) is consistent if $H_1(v) = H_1^*(v)$. The following lemma and theorem show that this definition of consistency, together with the connectivity and link availability conditions, are sufficient conditions to ensure full delivery in a self-pruning protocol.

Lemma 1 *If the global state G'' is connected and local states $G''(v)$ of all nodes v are consistent, then the forward node set determined by applying Wu and Dai's coverage condition on each node's local states form a CDS of G''.*

Proof: Suppose F is the forward node set derived using Wu and Dai's coverage condition on G'', F is a CDS of G'' (connectivity condition). Let F' be the forward node set derived using Wu and Dai's condition on local state $G''(v)$ of each node v. Consider each non-forward node $v \notin F'$ based v's local state, any two neighbors in $H_1(v)$ are connected via a replacement path in $G''(v)$.

17

Since $G''(v)$ is a subgraph of G'' and $H_1(v) = H_1^*(v)$ (consistency condition), any two neighbors in $H_1^*(v)$ are also connected by a replacement path in G''. Therefore, $v \notin F$. That is, F' is a superset of F and a CDS of G''. $\qquad \square$

Theorem 3 *If the connectivity, link availability, and consistency conditions are all satisfied, then Wu and Dai's generic self-pruning protocol ensures full coverage.*

Proof: Based on the connectivity and consistency conditions, the global state is connected and consistent. From Lemma 1, the corresponding forward node set F' is a CDS of the global broadcast state. Based on the link availability condition, each link in the global broadcast state corresponds to a valid link in the physical network. Therefore, all nodes in the physical network finally receive the broadcast. $\qquad \square$

Next we propose to use the *aggregated local view* to address the inconsistency problem. The inconsistency in the above example happens when node y removes link (y, w) in its local view before node x does so. As shown by the broadcast state 1 in Figure 9 (a), a broken link is first detected by the end nodes (1-hop neighbors). This link is not removed from local views of other nodes until the link failure is advertised via "Hello" messages. When k-hop information is used, it takes up to k "Hello" intervals for all affected nodes to update their local views. The solution is that once *a node advertises its 1-hop neighbor set, it cannot back away from it immediately*. Each node v keeps k recent versions of its local view, and uses the *aggregation* of those local views to make the forwarding/non-forwarding decision.

More formally, let $G_k^0(v) = (N_k(v), E_k(v))$ be the current local view of v, and $G_k^{-i}(v)$ be the local view i "Hello" intervals ago (or simply called view $-i$). Let $L_j(v) = (H_{j-1}(v) \times H_j(v)) \cap E_k(v)$ contains links between v's $(j-1)$- and j-hop neighbors. Links in $L_j(v)$ have an "age" of j; i.e., this information takes j rounds of "Hello" exchanges to build up. Also, $L_j^{-i}(v)$ corresponds to the $L_j(v)$ in view $-i$, which has an age of $i + j$. The aggregated local view of v, $G_k'(v)$, is defined as follows:

$$G_k'(v) = (N_k(v), L_1'(v) \cup L_2'(v) \cup ... \cup L_k'(v))$$

where

$$L_j'(v) = L_j(v) \cup L_j^{-1}(v) \cup ... \cup L_j^{-(k-j)}(v)$$

for $1 \leq j \leq k$. Note that the maximal age of $L_j'(v)$ is k.

In aggregated local views, the life span of each link is extended to compensate for the propagation delay of the link failure information. Aggregated local views are consistent; that is, local
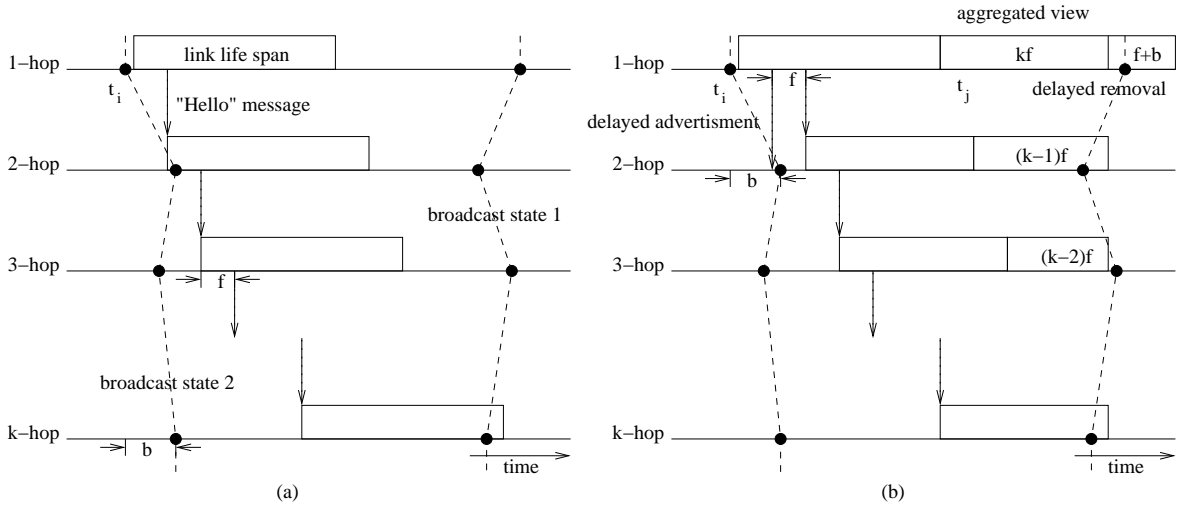
Figure 9: Life spans of a single link in local views of neighboring nodes. (a) Inconsistent local states may be sampled due to inconsistent local views and/or the broadcast delay. (b) Local states sampled from aggregation local views enhanced by delayed advertisement and delayed removal.

broadcast states sampled at the same physical time would be consistent (e.g., when the broadcast process completes instantly). Figure 8 (c) shows the aggregated local view of node y. Based on this local view, node y will still forward. Intuitively, once a node v appears as a neighbor of u (in the range of r_1) during the recent k intervals, it still has to be treated as a neighbor even if it currently moves out of u's visible range, but is still in u's effective neighbor set (as shown in Theorem 2).

Another cause of inconsistency is the backoff delay, where a new link is detected during a broadcast process. As shown in Figure 6, node u is initially a neighbor of w and later moves to x as its neighbor. If a "Hello" message is sent from x to w after x has made its decision, but before w's decision is made, then w's decision is made based upon information that is not available to x when it made its decision. Consider the following sequence of events as shown in Figure 6: (1) x decides its non-forwarding status, (2) u is detected by x as a new neighbor, (3) x advertises its new neighbor set, and (4) w believes that u is covered by x and becomes a non-forward node. In this case, u will never receive the broadcast packet. This situation is illustrated in Figure 9 (a) as broadcast state 2, where life spans of a single link in local views of neighboring nodes $(1, 2, \ldots, k$ hops away from the link) are shown. The new link is detected after a 1-hop neighbor has made its decision, and is propagated to a 2-hop neighbor before a decision is made there. One solution is the *delayed advertisement*. As shown in Figure 9 (b), a 1-hop neighbor will hold the advertisement of a new link if its age is younger than b, the maximal time period of a broadcast process. Thus a

19

new link detected during a broadcast process will not affect the decision of other nodes.

Two other problems may cause inconsistency in aggregated local views. The first one is illustrated by the broadcast state 1 in Figure 9 (a). When the broadcast process begins at a k-hop neighbor and ends at a 1-hop neighbor of a broken link, this link may be removed from the 1-hop neighbor's local view before the local state is sampled. Another problem is that a link failure may be detected by two end nodes in different times, a time skew bounded by the "Hello" interval f. Both problems can be solved by the *delayed removal* of a link from local views of its end nodes with a delay of $f + b$. Figure 9 (b) shows the effect of delayed advertisement, aggregated view, and delayed removal. After using delayed advertisement, if a link does not appear in the local view of a 1-hop neighbor, it will not appear in a i-hop ($i \geq 2$) neighbor's local state either. After using the aggregated view and delayed removal, the removal of a link from the local views of its 1-hop neighbors is extended by $kf + f + b$. Because this link will be removed from local views of all i-hop neighbors within k "Hello" intervals, the local views are always consistent. In addition, the time skew among two 1-hop nodes and the broadcast delay are compensated by the additional delay of $f + b$. Therefore, if a link appears in the local state of an i-hop neighbor, it will appear in both 1-hop neighbors' local views too. The following theorem can be easily proved based on the above reasoning.

Theorem 4 *If each node maintains its aggregated local view with delayed advertisement and delayed removal, then all local broadcast states sampled from these local views are consistent.*

3.3   Implementation details

According to Theorem 1, full coverage is guaranteed only when the network is sufficiently dense (i.e., connected with transmission range $r_1 - \Delta'$). In the following, we propose a mechanism, called *dual neighbor sets*, to relax the connectivity requirement under the cost of pruning efficiency. In sparse networks, using a small "Hello" transmission range may cause partition in the logical network. As shown in Figure 10 (a), when the "Hello" transmission range is $r_1 = r_2 - \Delta''$, neither node u nor v view node w as a neighbor, because they cannot receive "Hello" messages from node w. Therefore, both nodes u and v are non-forward nodes, and no one will forward the broadcast packet to node w. On the other hand, using a larger r_1 violates the link availability condition in Theorem 2. The broadcast process may fail due to the lack of a buffer zone. Figure 10 (b) shows the situation when $r_1 = r_2$ is used. Node u is a non-forward node, and node v is supposed to

(a) "Hello" range r1 causes partition   (b) "Hello" range r2 causes link failure
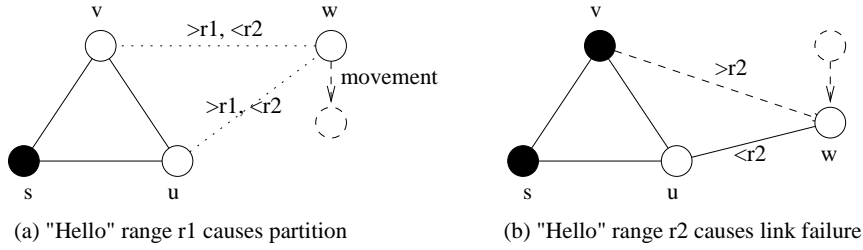
Figure 10: A network with one mobile node w (a) before the movement and (b) after the movement. Dotted lines represent undetected physical links. The dashed line represents an undetected broken link.

forward the broadcast packet to node w. However, node w moves out of the transmission range of v and never receives this packet.

Our solution is based on maintaining two neighbor sets. The *effective neighbor set* consists of all nodes within the actual transmission r_2, and the *advertised neighbor set* consists of only nodes with distance less than r_1. If v is a non-forward node, every pair of nodes in v's effective neighbor set must be connected via a replacement path. In this case, node v in Figure 10 (a) views w as a neighbor and becomes a forward node. On the other hand, only the advertised neighbor set is propagated to its k-hop neighbors. Because link (v, w) in Figure 10 (b) is invisible to node u, node u also forwards the broadcast packet and ensures the coverage. Note that this method is conservative. If link (v, w) is still available, making node u a forward node causes extra redundancy. Using the dual neighbor set mechanism, the connectivity condition in Theorem 1 is relaxed without violating the link availability condition. It is sufficient that the physical network is connected with transmission range $r_2 - \Delta'$.

If the physical distance among two neighbor can be estimated based on the signal strength of a received message, the dual neighbor sets can be implemented via sending "Hello" messages with transmission range r_2. Each node classifies its neighbors into effective and advertised neighbors based on their estimated distances. When the signal strength is unavailable or inaccurate, each node will send two types of "Hello" messages. The first type of "Hello" messages, sent via transmission range r_2, are used to detect effective neighbors. The second, sent via transmission range r_1, are used to identify advertised neighbors.
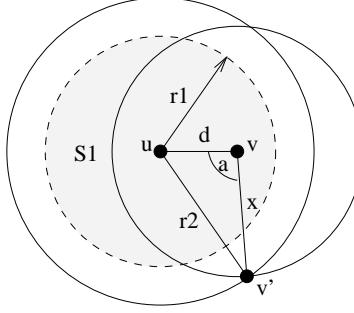
21

Figure 11: Calculation of the probability that a neighbor within the "Hello" transmission range (r_1) moves out of the normal transmission range (r_2).

3.4 Analytical study

Based on Theorem 2, in order to guarantee that a neighbor (within r_1) at t_0 is an effective neighbor (within r_2) at a time $t_1 = t_0 + f$, r_1 must be smaller than $r_2 - 2sf$ for a given maximal node speed s and time period f. In this section, we show that the probability, p, that a node within r_1 at t_0 moves out of range r_2 at t_1 is reasonably small with a much larger r_1. We assume a mobility model similar to the random direction model [21], where each node is moving at a random speed in $[0, s]$ to a random direction in $[0, 2\pi]$. This is a simplified model for ease of probabilistic analysis. In addition, this model usually represents the worst case in terms of relative distance between two nodes in a given interval.

Consider two neighboring nodes u and v as shown in Figure 11. Node v is within u's "Hello" transmission range (the shadowed area) at time t_0, and moves to position v' at t_1. Assume that their distance at t_0 is d, and v moves a distance of x with respect to u at t_1. The probability that v moves out of the actual transmission range of u is

$$
p(x, d) = \begin{cases} 0 & : \quad x < r_2 - d \\ 1 - \frac{\alpha}{\pi} & : \quad r_2 - d \leq x \leq r_2 + d \\ 1 & : \quad x > r_2 + d \end{cases} \tag{1}
$$

where

$$
\alpha = \cos^{-1}\left(\frac{x^2 + d^2 - r_2^2}{2dx}\right)
$$

is the largest value of $\angle uvv'$ that satisfies $d(u, v') \leq r_2$. The probability that *any* node within the "Hello" transmission range of u moves out of its actual transmission range at t_1 is

$$
p(x) = \int_0^{r_1} \frac{2\pi d}{S_1} p(x, d) \, \mathrm{d}d = \int_0^{r_1} \frac{2d}{r_1^2} p(x, d) \, \mathrm{d}d \tag{2}
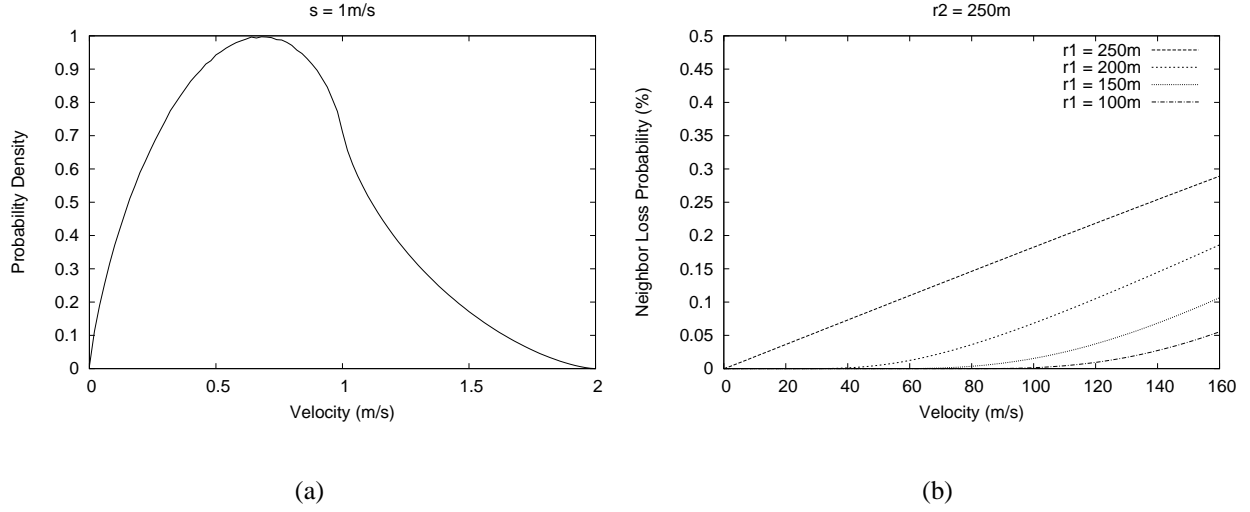$$

22

Figure 12: The probability function $f_{|\vec{V}|}(t)$ of the random joint mobility vector (a), and the probability that a neighbor within a given "Hello" transmission range r_1 moves out of the actual transmission range r_2 (b).

where

$$S_1 = \pi r_1^2$$

is the area within the "Hello" transmission range. The probability that a node with *any* constant relative speed t with respect to u moves out of the actual transmission range is

$$p = \int_0^{2s} f_{|\vec{V}|}(t)\, p(ft)\, \mathrm{d}t \tag{3}$$

Here $\vec{V} = \vec{V_v} - \vec{V_u}$ is the random joint mobility vector between any two mobile nodes u and v, where $\vec{V_u}$ ($\vec{V_v}$) is the random mobility vector of node u (v). Note that equation (1) still holds, as the direction of $\vec{V}$ is also uniformly distributed in $[0, 2\pi]$, and is independent of the speed of $\vec{V}$, $|\vec{V}|$. We know that $|\vec{V}|$ is between 0 and $2s$; $|\vec{V}| = 0$ when $\vec{V_u} = \vec{V_v}$, and $|\vec{V}| = 2s$ when $\vec{V_u} = -\vec{V_v}$ and $|\vec{V_u}| = |\vec{V_v}| = s$. However, its probability function, $f_{|\vec{V}|}(t)$, is unknown. McDonald and Znati [18] conducted a probabilistic analysis on the joint mobility of two nodes, but their analysis is based on the random walk mobility model [4], where the mobility vector of each node is the sum of several epochs; each epoch has different speed, direction, and duration. Li, Hou and Sha's analysis [17] is based on the same mobility model as ours, but their analysis is simplified by the implicit assumption that node u is fixed and $|\vec{V}|$ is uniformly distributed in $[0, s]$. Here we calculate $f_{|\vec{V}|}(t)$

23

at a given t as

$$
\begin{aligned}
f_{|\vec{V}|}(t) &\approx \frac{F_{|\vec{V}|}(t + \delta t) - F_{|\vec{V}|}(t)}{\delta t} \\
&= \frac{P(t \leq |\vec{V}| \leq t + \delta t)}{\delta t} \\
&= \oint_{(0,0)}^{(2\pi,s)} \oint_{(0,0)}^{(2\pi,s)} \frac{R(\vec{V_u}, \vec{V_v}, t, t + \delta t)}{(2\pi s)^2 \delta t} \mathrm{d}\vec{V_u} \mathrm{d}\vec{V_v}
\end{aligned}
\tag{4}
$$

where $F_{|\vec{V}|}(t)$ is the distribution function, δt is a small positive value, and

$$
R(\vec{V_u}, \vec{V_v}, a, b) = \begin{cases} 1 & : \quad a \leq |\vec{V_v} - \vec{V_u}| \leq b \\ 0 & : \quad \text{otherwise} \end{cases}
$$

Figure 12 (a) shows the distribution of $|\vec{V}|$ calculated from (4), when $s = 1m/s$ and $\delta t = 0.001m/s$. Note that the probability that $|\vec{V}| > 1.5s$ is small ($\leq 5\%$). Based on this distribution, we calculate the probability p that any node within the "Hello" transmission range ($r_1 = 100$, 150, 200, and 250) of u moves out of its actual transmission range ($r_2 = 250m$) during a "Hello" interval ($f = 1s$), when the maximal single node speed s varies from 0 to $160m/s$. As shown in Figure 12 (b), we can use an r_1 that is much larger than $r_2 - 2sf$, and still expect a low probability that an effective neighbor moves out of the actual transmission range. For example, when $r_1 = 200m$ and $s = 80m/s$, the probability of losing an effective neighbor is less than 5%. Note that the corresponding r_1 that guarantees the availability of link (u, v) at time t_1 is $r_2 - 2sf = 90m$. When $r_1 = 100m$ and $s = 160m/s$, the probability of losing an effective neighbor is about the same. On the other hand, there is no r_1 that can guarantee the link availability, as $2sf = 320m > r_2$.

4 Simulation

Simulations are conducted to evaluate the proposed method and explore appropriate "Hello" transmission ranges that achieve high delivery ratio with low broadcast redundancy under various mobility levels. We also evaluate the effectiveness of two implementation options that use dual neighbor sets to improve the delivery ratio under various environments.

Table 1: Simulation parameters

Parameter	Value
Network area	$900 \times 900 \ m^2$
Actual transmission range	$250 \ m$
Number of nodes	50, 100
Average moving speed	1-160 m/s
"Hello" interval	0.75-1.25 s
Priority type	node id
Simulation time	100 s
Number of trials	20
Confidence level	95%

4.1 Simulation environment

The proposed mobility management method is simulated on *ns-2(1b7a)* [9] and its CMU wireless extension. We extend Wu and Dai's coverage condition by using two transmission ranges r_1 (for "Hello" messages) and r_2 (for actual transmission). When $r_1 = r_2$, the new algorithm is equivalent to the original generic self-pruning protocol (Generic in Figure 4). We also simulate the dual neighbor sets enhancement for sparse networks. **The consistency enforcement mechanisms (i.e., aggregated local views, delayed advertisement, and delayed removal) are not implemented.**

The broadcast traffic rate is 10 packet per second and 64 bytes per packet. Each packet is issued from a randomly selected node. Since our purpose is to observe the behavior of self-pruning protocols under mobile environments, all simulations use an ideal MAC layer without contention or collision. If a node sends a packet, all neighbors within its transmission range will receive this packet after a short propagation delay. We assume that accurate location information is either unavailable, or unable to predict the existence of wireless links due to the irregular variation of transmission range. It was shown in [7] that the contribution of a backoff delay to the protocol efficiency is trivial except for SBA. Therefore, our implementation of the proposed method does not use a backoff delay.

The mobility model used in the simulation is the random direction model [21]. In this model, each node heads in a random direction and moves at a random speed until it reaches the boundary of the area, where it selects new direction and speed and keeps moving. Our mobility pattern generator is from [4], which has a parameter called average moving speed (V_{avg}). For a given V_{avg}, the
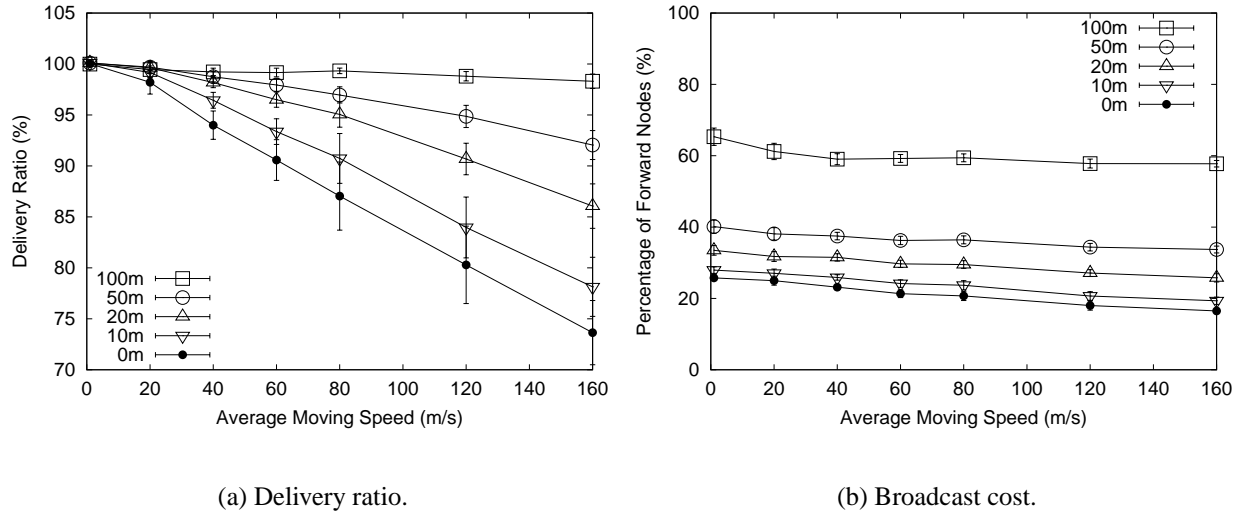
(a) Delivery ratio.   (b) Broadcast cost.

Figure 13: Performance of the original single neighbor set method in relatively dense (100 nodes) networks.

speed of each node is randomly selected from the range $[0, 2V_{avg}]$. Note that the random direction model usually yields sparser networks and higher mobility than the commonly used random way-point model [4]. Therefore, a reliable protocol in this simulation study is a reliable protocol under the random waypoint model, but not vice versa. Other simulation parameters are listed in Table 1.

4.2 Simulation results

Figure 13 shows simulation results in relatively dense networks (100 nodes), with buffer zone width (i.e., $r_2 - r_1$) varying from $0m$ to $100m$. As expected, high delivery ratio ($\geq 98\%$) can be achieved with large buffer zone width ($100m$) in highly mobile networks (with average speed $160m/s$). Note that this buffer zone width is much smaller than that required by Theorem 2, which should be at least $2 \times 2V_{avg} = 640m$. This also confirms our prediction in Section 3.4 that high delivery ratio can be achieved using a small buffer zone. The only problem is the high broadcast cost ($\geq 60\%$ forward nodes). If the network mobility level is known, we can select the buffer zone width based on the mobility level to balance the delivery ratio and redundancy. For example, at average speed $120m/s$, we can use a buffer zone width of $50m$, which achieves 95% delivery ratio with 40% forward nodes. At average speed $40m/s$, a $10m$ buffer zone achieves the same delivery ratio with only 30% forward nodes.

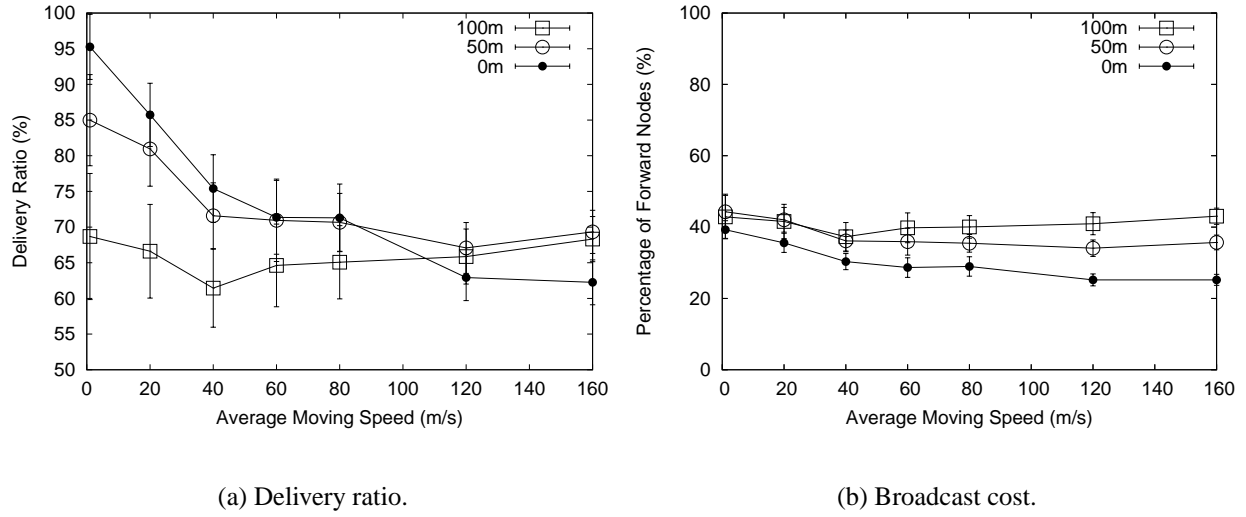(a) Delivery ratio.   (b) Broadcast cost.

Figure 14: Performance of the original single neighbor set method in relatively sparse (50 nodes) networks.

Figure 14 shows the performance of the proposed method in relatively sparse networks (50 nodes). When a $0m$ buffer zone is used, the delivery ratio drops rapidly as the average speed increases. Using a larger buffer zone width ($50m$ or $100m$) improves the delivery ratio under high mobility level, but performs poorly under low mobility level. The delivery ratio is low (85% and 70%), even with trivial mobility ($1m/s$). One reason for the low delivery ratio in sparse networks is the relatively low redundancy. The broadcast cost is low (40%) under different buffer zone widths. This is because the broadcast process fails in most cases, where only a few nodes receive and forward the packet.

Simulation results in [7] showed that all self-pruning protocols have lower delivery ratio in sparse networks than in dense networks under the same mobility level. Another reason is that when the network is not dense enough, the connectivity requirement in Theorem 1 is not satisfied, and therefore, cannot guarantee the coverage. This problem can be solved with the dual neighbor set enhancement introduced in Section 3.3. Figure 15 shows the performance of the enhanced scheme, where all neighbors within the actual transmission range r_2 are put into the effective neighbor set, and only neighbors within transmission range r_1 are put into the advertised neighbor set. With this enhancement, high delivery ratio ($\geq 90\%$) can still be achieved under the highest mobility level. However, the corresponding broadcast cost is also higher (80%). Under low and median mobility ($V_{avg} \leq 40$), a smaller buffer zone width ($50m$) can be used to reduce the broadcast cost to 60%.
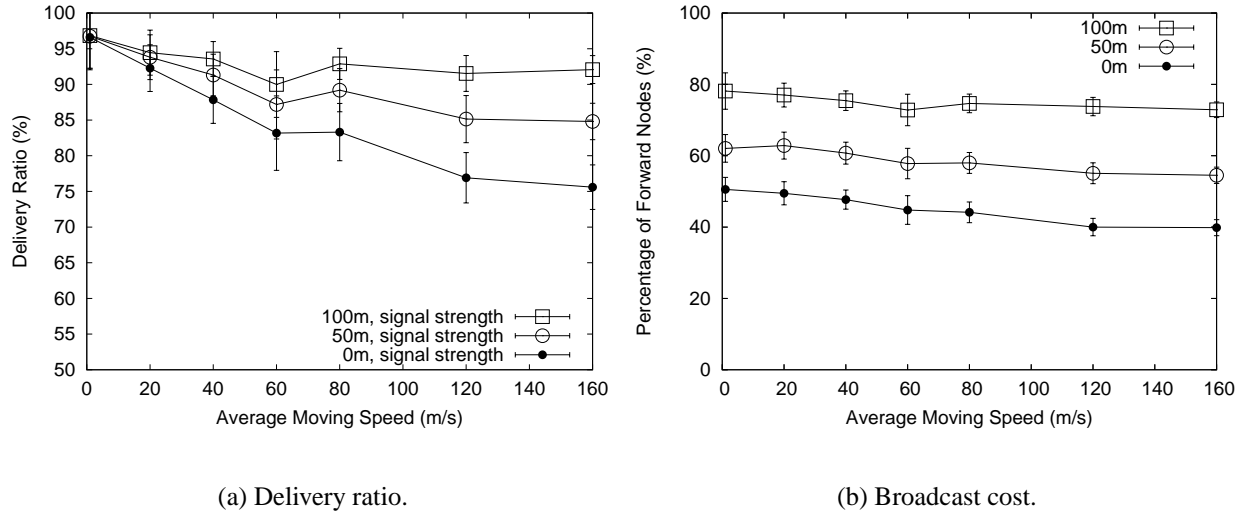
27

(a) Delivery ratio.

(b) Broadcast cost.

Figure 15: Performance of the dual neighbor set enhancement in relatively sparse (50 nodes) networks.

Overall, simulation results show that balance between delivery ratio and broadcast redundancy can be achieved by adjusting the buffer zone width based on the network mobility level. As predicted by our probabilistic analysis, for each mobility level, high delivery ratio can be achieved with a buffer zone much thinner than required by Theorem 2. The dual neighbor set enhancement is proved successful in relaxing the connectivity requirement in Theorem 1, and achieves high delivery ratio in sparse networks.

5 Conclusions

In this paper, we have proposed a mobility management method based on the use of two transmission ranges. Using this mechanism, we have also extended Wu and Dai's coverage condition to a dynamic environment where network topology is allowed to change, even during the broadcast process. In addition, connectivity, link availability, and consistency issues related to neighborhood information of different nodes have also been addressed. The proposed scheme can also be extended to provide mobility management for other activities such as topology control in MANETs [32].

The constraint used on $r_2 - r_1$ in this paper is conservative. Our probabilistic analysis suggests

that high delivery ratio can still be achieved with a larger r_1. Simulation results show that the proposed method and the dual neighbor set enhancement achieve good balance between delivery ratio and broadcast redundancy by adjusting the value of r_1 based on the network mobility level. A future extension would be automatic buffer zone width adjustment that adapts to the neighborhood mobility level.

In Wu and Dai's coverage condition, node id is used to break a tie. We could also use the notion of *relative mobility* [1], defined as absolute relative speed averaged over time, for tie breaking. In general, a node with high relative mobility is more prone to unstable behavior than a node with less relative mobility and therefore should be pruned (from being a forward node) when possible. In this case, relative mobility is calculated locally through some form of approximation and distributed through piggybacking with regular "Hello" messages. Our future work also includes the adoption of vast results from the distributed system community with regards to global/local states and view consistency [24].

References

[1] B. An and S. Papavassiliou. A mobility-based clustering approach to support mobility management and multicast routing in mobile ad-hoc wireless networks. *International Journal of Network Management*, 11(6):387–395, Nov.-Dec. 2001.

[2] F. Bai, N. Sadagopan, and A. Helmy. IMPORTANT: a framework to systematically analyze the impact of mobility on performance of routing protocols for ad-hoc networks. In *Proc. of IEEE Infocom*, Mar./Apr. 2003.

[3] S. Basagni, I. Chlamtac, A. Farago, V. R. Syrotiuk, and R. Talebi. Route selection in mobile multimedia ad hoc networks. In *Proc. of IEEE MOMUC*, Nov. 1999.

[4] T. Camp, J. Boleng, and V. Davies. A survey of mobility models for ad hoc network research. *Wireless Communication & Mobile Computing (WCMC): Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications*, 2(5):483–502, 2002.

[5] B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris. Span: An energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks. In *Proc. of MobiCom*, pages 85–96, July 2001.

[6] F. Dai and J. Wu. An extended localized algorithm for connected dominating set formation in ad hoc wireless networks. *IEEE Transaction on Parallel and Distributed Systems*, Oct. 2004. accepted to appear.

[7] F. Dai and J. Wu. Performance comparison of broadcast protocols for ad hoc networks based on self-pruning. *IEEE Transaction on Parallel and Distributed Systems*, Nov. 2004. accepted to appear.

[8] R. Dube, C. Rais, K. Wang, and S. Tripathi. Signal stability based adaptive routing (SSA) for ad hoc mobile networks. *IEEE Personal Communications*, 4(1):36–45, Feb. 1997.

[9] K. Fall and K. Varadhan. The *ns* manual. The VINT Project, UCB, LBL, USC/ISI and Xerox PARC, http://www.isi.edu/nsnam/ns/doc/, Apr. 2002.

[10] M. Grossglauser and D. Tse. Mobility increases the capacity of ad-hoc wireless networks. In *Proc. of IEEE Infocom*, Apr. 2001.

[11] P. Gupta and P. R. Kumar. The capacity of wireless networks. *IEEE Transactions on Information Theory*, 46(2):388404, Mar. 2000.

[12] Z. J. Haas, J. Y. Halpern, and L. Li. Gossip-based ad hoc routing. In *Proc. of IEEE Infocom*, June 2002.

[13] Z. J. Haas and M. R. Pearlman. The zone routing protocol (ZRP) for ad hoc networks. IETF Internet Draft (draft-zone-routing-protocol-00.txt), Nov. 1997.

[14] D. B. Johnson, D. A. Maltz, Y.-C. Hu, and J. G. Jetcheva. The dynamic source routing protocol for mobile ad hoc networks. IETF Internet Draft (draft-ietf-manet-dsr-05.txt), Mar. 2001.

[15] W.-I. Kim, D.-H. Kwon, and Y.-J. Suh. A reliable route selection algorithm using global positioning systems in mobile ad-hoc networks. In *Proc. of ICC*, June 2001.

[16] Y.-B. Ko and N.H. Vaidya. Location-aided routing (LAR) in mobile ad hoc networks. In *Proc. of ACM/IEEE MobiCom*, Oct. 1998.

[17] N. Li, J. C. Hou, and L. Sha. Design and analysis of an MST-based topology control algorithm. In *Proc. IEEE Infocom*, Mar./Apr. 2003.

[18] A. B. McDonald and T. F. Znati. A mobility-based framework for adaptive clustering in wireless ad hoc networks. *IEEE JSAC, Special Issue on Ad-Hoc Networks*, Aug. 1999.

[19] W. Peng and X. Lu. On the reduction of broadcast redundancy in mobile ad hoc networks. In *Proc. of MobiHoc*, pages 129–130, 2000.

[20] C. E. Perkins, E. M. Royer, and S. R. Das. Ad hoc on-demand distance vector (AODV) routing. IETF Internet Draft (draft-ietf-manet-aodv-08.txt), Mar. 2001.

[21] E. Royer, P. Melliar-Smith, and L. Moser. An analysis of the optimum node density for ad hoc mobile networks. In *Proc. of ICC*, 2001.

[22] N. Sadagopan, F. Bai, B. Krishnamachari, and A. Helmy. Paths: analysis of path duration statistics and their impact on reactive MANET routing protocols. In *Proc. of MobiHoc*, June 2003.

[23] C. Schindelhauer, T. Lukovszki, S. Rhrup, and K. Volbert. Worst case mobility in ad hoc networks. In *Proc. of the 15th annual ACM symposium on Parallel algorithms and architectures*, Aug. 2003.

[24] M. Singhal and N. G. Shivaratri. *Advanced Concepts in Operating Systems: Distributed, Database, and Multiprocessors Operating Systems*. McGraw-Hill, Inc., 1994.

[25] I. Stojmenovic, M. Seddigh, and J. Zunic. Dominating sets and neighbor elimination based broadcasting algorithms in wireless networks. *IEEE Transactions on Parallel and Distributed Systems*, 13(1):14–25, Jan. 2002.

[26] W. Su, S.-J. Lee, and M. Gerla. Mobility prediction and routing in ad hoc wireless networks. *International Journal of Network Management*, 11(1):3–30, Jan.-Feb. 2001.

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

[28] C.-K. Toh. Associativity-based routing for ad-hoc mobile networks. *IEEE Personal communications*, 4(2):36–45, Mar. 1997.

[29] Y.-C. Tseng, S.-Y. Ni, Y.-S. Chen, and J.-P. Sheu. The broadcast storm problem in a mobile ad hoc network. *Wireless Networks*, 8(2/3):153–167, Mar.-May 2002.

[30] Y.-C. Tseng, S.-Y. Ni, and E.-Y. Shih. Adaptive approaches to relieving broadcast storms in a wireless multihop mobile ad hoc network. *IEEE Transactions on Computers*, 52(5):545–557, May 2003.

[31] J. Wu and F. Dai. Broadcasting in ad hoc networks based on self-pruning. In *Proc. of IEEE Infocom*, Mar./Apr. 2003.

[32] J. Wu and F. Dai. Mobility-sensitive topology control in mobile ad hoc networks. In *Proc. of IPDPS*, May 2004.

[33] J. Wu and H. Li. On calculating connected dominating set for efficient routing in ad hoc wireless networks. In *Proc. of DiaLM*, pages 7–14, 1999.

Appendix: special cases of the coverage condition

Wu and Li's algorithm (static): Wu and Li [33] proposed a *marking process* to determine a set of *gateways* (i.e., forward nodes) that form a CDS: a node is marked as a gateway if it has two neighbors that are not directly connected. Two pruning rules are used to reduce the size of the resultant CDS. According to pruning Rule 1, a gateway v can become a non-gateway if all of its neighbors are also neighbors of another node u that has higher priority value; that is, v's neighbor set is *covered* by u and u is called a *coverage node*. According to pruning Rule 2, a marked node can be unmarked if all of its neighbor set is covered by two other nodes that are directly connected and have higher priority values.

Dai and Wu's algorithm (static): Dai and Wu [6] extended the previous algorithm by using a more general pruning rule called Rule k: a gateway becomes a non-gateway if all of its neighbors are also neighbors of any one of k other nodes that are connected and have higher priority values. Rules 1 and 2 are special cases of Rule k where k is restricted to 1 and 2, respectively.

Span (static): Chen, Jamieson, Balakrishman, and Morris [5] proposed the *Span* protocol to construct a set of forward nodes (called *coordinators*). A node becomes a coordinator if it has two neighbors that are not directly connected, indirectly connected via one intermediate coordinator, or indirectly connected via two intermediate coordinators. Before a node changes its status from non-coordinator to coordinator, it waits for a backoff delay which is computed from its energy level, node degree, and the number of pairs of its neighbors that are not directly connected. The

(a) Marking process   (b) Rule 1   (c) Rule 2   (d) Rule k

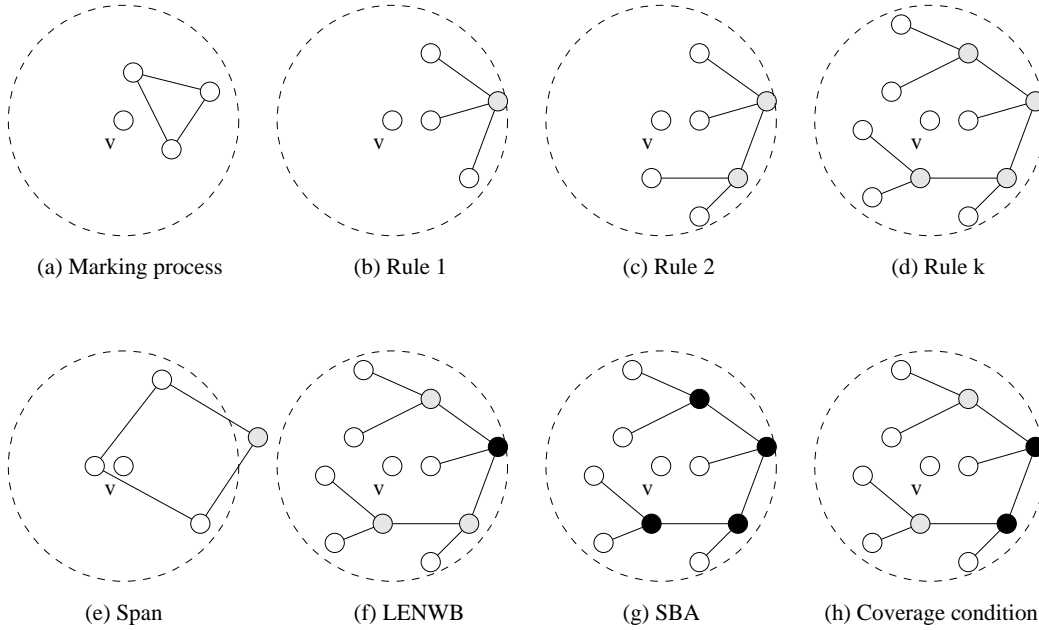(e) Span   (f) LENWB   (g) SBA   (h) Coverage condition

Figure 16: Node v in the center of each subgraph can be self-pruned by the corresponding protocol. Nodes in the transmission range of node v (the dashed circle) are neighbors of v. Gray nodes have higher priorities (e.g., higher id's) than v. Black nodes are visited nodes that have forwarded the broadcast packet.

backoff delay can be viewed as a priority value, such that nodes with shorter backoff delay have a higher chance of becoming coordinators.

LENWB (dynamic): Sucec and Marsic [27] proposed the *Lightweight and Efficient Network-Wide Broadcast* (LENWB) protocol, which computes the forward node status on-the-fly. Whenever node v receives a broadcast packet from a neighbor u, it computes the set C of nodes that are connected to u via nodes that have higher priority values than v. If v's neighbor set, $N(v)$ (i.e., $N_1(v)$) is contained in C, node v is a non-forward node; otherwise, it is a forward node.

SBA (dynamic): Peng and Lu [19] proposed the Scalable Broadcast Algorithm (SBA) to reduce the number of forward nodes. As in LENWB, the status of a forward node is computed on-the-fly. When a node v receives a broadcast packet, instead of forwarding it immediately, v will wait for a backoff delay. For each neighbor u that has forwarded the broadcast packet, node v removes $N(u)$ from $N(v)$. If $N(v)$ does not become empty after the backoff delay, node v becomes a forward node; otherwise, node v is a non-forward node.

Stojmenovic's algorithm (static/dynamic): Stojmenovic, Seddigh, and Zuinic [25] extended Wu and Li's algorithm in two ways: (1) Suppose every node knows its accurate geographic position, only 1-hop information is needed to implement the marking process and Rules 1 and 2. That is, each node only maintains a list of its neighbors and their geographic positions (connections among neighbors can be derived). (2) The number of forward nodes are further reduced by a neighbor elimination algorithm similar to the one used in SBA.

The difference among above special cases is illustrated by Figure 16. To have a fair comparison, each node is equipped with only 2-hop information. That is, all coverage nodes are 2 hops away from the covered node. Node v in subgraphs (a), (b), and (c) can be pruned by Wu and Li's algorithm. Node v in subgraphs (a) to (d) can be pruned by Dai and Wu's algorithm. Node v in subgraphs (a), (b), (c), and (e) can be pruned by Span. Node v in subgraphs (a) to (f) can be pruned by LENWB. Node v in subgraphs (a) and (g) can be pruned by SBA. Node v in subgraphs (a), (b), (c), and (g) can be pruned by Stojmenovic's algorithm. Node v in all subgraphs can be pruned by the coverage condition. A static protocol has only gray nodes, whereas a dynamic protocol has both gray and black nodes.