An O(log n) Dominating Set Protocol for Wireless Ad-Hoc Networks under the Physical Interference Model[*]

Christian Scheideler[†]
Institut für Informatik
Technische Universität
München, Germany
scheideler@in.tum.de

Andrea W. Richa
Dept. of Compute Science
and Engineering
Arizona State University, USA
aricha@asu.edu

Paolo Santi
Instituto di Informatica e
Telematica del CNR
Pisa, Italy
paolo.santi@iit.cnr.it

ABSTRACT

Dealing with interference is one of the primary challenges to solve in the design of protocols for wireless ad-hoc networks. Most of the work in the literature assumes localized or hop-based interference models in which the effect of interference is neglected beyond a certain range from the transmitter. However, interference is a more complex phenomenon that cannot, in general, be captured by localized models, implying that protocols based on such models are not guaranteed to work in practice. This paper is the first to present and rigorously analyze a distributed dominating set protocol for wireless ad-hoc networks with $O(1)$ approximation bound based on the physical interference model, which accounts for interference generated by all nodes in the network. The proposed protocol is fully distributed, randomized, and extensively uses physical carrier sensing to reduce message overhead. It does not need node identifiers or any kind of prior information about the system, and all messages are of constant size (in bits). We prove that, by appropriately choosing the threshold for physical carrier sensing, the protocol stabilizes within a logarithmic number of communication rounds, w.h.p., which is faster than the runtime of any known distributed protocol without prior knowledge about the system under any wireless model that does not abstract away collisions.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Wireless Networks

General Terms

Algorithms, Theory

Keywords

Ad hoc networks, dominating set, physical interference model, self-stabilization

[*]This work originated from a discussion at the Dagstuhl Seminar on Geometry in Sensor Networks, 2007.

[†]Supported by DFG grant SCHE 1592/1-1.

MobiHoc'08, May 26–30, 2008, Hong Kong SAR, China.

1. INTRODUCTION

A major challenge in the design of protocols for wireless multi-hop networks is related to modeling complex physical phenomena such as radio wave propagation and interference. On one hand, using oversimplified radio propagation/interference models leads to the design of protocols that, although efficient in terms of computational complexity and message overhead, might display a considerably different behavior from what is expected when utilized in a practical scenario. On the other hand, using very complex radio propagation/interference models might hinder the design and analysis of efficient protocols. Hence, finding the adequate compromise between model accuracy and computational/communication efficiency is at the heart of the successful design of protocols for wireless multi-hop networks.

The model accuracy/efficiency tradeoff has not been adequately addressed so far. Computer scientists have mostly focused their attention on designing distributed protocols for wireless multi-hop networks based on simplistic models, such as the unit disk graph model (UDG) for radio wave propagation [17], and graph-based models for interference [29]. Classical graph-related problems such as distributed coloring, dominating set construction, clustering, etc., have been successfully addressed in the past. However, these protocols are guaranteed to work only if the models' assumptions are fulfilled, which is hardly the case in practice. In particular, the UDG model assumes isotropic propagation of the radio signal with distance, which is unlikely to happen in the real world due to phenomena such as scattering, reflection, and diffraction of the radio signal. Concerning interference, graph-based models assume that only nodes within d hops from a certain receiver u can interfere with u, where d is a small constant. A major shortcoming of these models is that interference is modeled as a localized phenomenon, which is not the case in practice. It is well known, in fact, that a transmitter can corrupt message reception even at very large distances, especially in large networks with several concurrent transmissions. A similar criticism applies to another interference model commonly used in the design of protocols for wireless multi-hop networks, namely the protocol interference model [12]. In this model, a transmission from node v to node u is considered successful iff there is no other transmitter within a certain (constant) range centered at u. On the other hand, the communication engineering community has focused on deriving accurate models for radio wave propagation and interference, but has devoted little effort to the design of protocols with *proven* performance guarantees based on these models.

In this paper, we try to bridge this gap by introducing a fully distributed, randomized protocol called TWIN for building a dominating set in logarithmic time, w.h.p.[1], in a wireless multi-hop network based on more realistic radio propagation and interference models with respect to those commonly used in the distributed algorithms community. In particular, we assume radio wave propagation obeys a cost-based generalization of the log-distance path loss model, which can be used to resemble a close approximation of the well-known log-normal shadowing model. For what concerns interference, we use the physical interference model introduced in [12], in which a message is correctly received if and only if the SINR value at the receiver is larger than a certain threshold.

A notable feature of the proposed algorithm is that, although it is a *localized* distributed protocol, it is guaranteed to work properly under a 'global' interference model such as the physical model. This is made possible by extensive use of physical carrier sensing. Another positive feature of using physical carrier sensing is reduction of the communication overhead: the protocol is shown to converge to a constant density dominating set (i.e., a dominating set with a constant number of nodes per unit square) within $O(\log n)$ communication rounds, w.h.p. Having such a dominating set is important because it can be used as a building block for designing more complex distributed protocols. For instance, in case of WSNs, once local leaders (nodes in the dominating set) are elected, network-layer functionalities such as broadcast and convergecast can be more easily implemented on top of them. Also application-layer functionalities such as data aggregation significantly benefit from the existence of a dominating set.

2. OUR CONTRIBUTIONS

We will focus on the dominating set problem. The classical dominating set problem is defined as follows. Given an undirected graph $G = (V, E)$, a subset $U \subseteq V$ is called a *dominating set* of G if all nodes $v \in V$ are either in U or have an edge to a node in U. The *density* of a dominating set is the maximum over all nodes $v \in U$ of the number of neighbors that v has in U. In our context, the neighborhood of a node v consists of all nodes within its transmission range, i.e., all nodes that can receive a message from v under the ideal case that only the background noise is there. (In simple wireless models like the unit disk graph model, this neighborhood consists of all nodes within a certain distance from v.) A constant density dominating set is a constant approximation of a minimum dominating set (i.e., a dominating set of minimum cardinality). The problem of finding a minimum dominating set has been shown to be NP-complete even when restricted to unit disk graphs [7] and, hence, approximation algorithms are of interest.

Our contributions are two-fold: *i*) we propose a new model for wireless communication based on the physical interference model which incorporates physical carrier sensing and which closely approximates the log-normal shadowing model; and *ii*) we demonstrate how to develop and analyze algorithms on top of this model by presenting a local-control algorithm for building a constant density dominating set, which we define below.

[1] "w.h.p." means with probability at least $1 - 1/n^c$ for any constant $c > 0$, where n is the number of network nodes.

A notable feature of our algorithm, which we call TWIN, is that nodes do not need to have *any* a priori knowledge about the other nodes, not even an estimate on their total number. Also, distinct identification numbers of any form are not required so that our protocol may even be applicable to the important field of sensor networks.

In order to obtain a constant density dominating set under an arbitrary distribution of nodes, our algorithm has two main, inter-dependent components: the *density estimation stage*, during which each node obtains an estimate on the number of nodes within its transmission range, and accordingly adjusts its transmission probability, and a *leader election stage*, which is done using the novel concept of "twin" nodes introduced in Section 5. The notion of twins forces leaders to wake up in a pair-wise fashion, so that they can check whether the density of nodes becoming leaders is too high. This ensures that the number of leader nodes within the transmission range of a node never exceeds a constant, so that the leader election process does not run into oscillation problems. Unfortunately, in some situations formation of twins cannot be enforced (such as isolated nodes), so we also need the concept of "single" leaders. However, our protocol privileges the "twin" status when electing leaders, and our strategy will be to form singles only as a last resort. Using these concepts, we can construct a simple protocol with the following performance.

THEOREM 2.1. *The TWIN protocol establishes a constant density dominating set in $O(\log n)$ communication rounds, w.h.p., where n is the number of nodes.*

We also discuss how to make this protocol self-stabilizing, i.e., to establish a constant density dominating set starting from any initial state. Note that a self-stabilizing protocol is able to tolerate dynamic network conditions due to, e.g., node failures, node mobility, and so on.

Interestingly, the runtime bound is only possible because our protocol uses physical carrier sensing. It is known that if physical carrier sensing is not available and nodes have no estimate of the size of the network, then it takes $\Omega(n)$ steps on expectation for a single message transmission to be successful [15] in any protocol. Also interesting is the fact that our $O(\log n)$ algorithm improves upon the best previous result in [19] under a much more restricted, bounded interference model.

3. RELATED WORK

Since interference is a major factor limiting performance in wireless multi-hop networks, a lot of effort has been invested in deriving realistic interference models, and to analyze network performance under such models. A seminal work in this area is [12], in which Gupta and Kumar study the transport capacity of wireless networks under two different interference models, the physical and the protocol interference model. Contrary to the physical model, which is used in this paper, the protocol interference model is a localized model, since decision on whether a certain communication is successful depends only on the presence of concurrent transmitters within a bounded area centered at the receiver. Another localized interference model commonly used in the literature is the graph-based interference model, in which a certain communication graph representing communication links is assumed, and only links whose endpoints are up to a certain hop distance d on the communication graph from link (u, v) can interfere with (u, v) [29].

Due to their simplicity and the fact that they somehow resemble the behavior of the 802.11 MAC layer, localized interference models have been mostly used in the literature to design interference-aware protocols. This is the case, for instance, for the protocols presented in [2, 21, 25, 26]. Given the complexity of dealing with physical interference, only a few protocols based on this model have been proposed so far. For example, [10, 11, 14] consider the problem of scheduling transmissions, but they provide solutions which are computationally infeasible even for a small size network. Only recently, a computationally efficient algorithm for scheduling transmissions under the physical interference model with a provable approximation bound has been proposed [4]. The physical interference model has been recently used in [9, 23] to study the complexity of scheduling a set of link demands in the shortest possible time and of one-shot scheduling (scheduling as many transmissions as possible in a single communication slot), which are both shown to be NP-complete for wireless networks in a 2-dimensional Euclidean space.

Various distributed algorithms have been proposed for finding good approximations of minimum dominating sets in arbitrary graphs (see, for example, [8, 19, 20, 22]). Alzoubi et al. [3] presented the first constant approximation algorithm for the minimum connected dominating set problem in unit-disk graphs with $O(n)$ and $O(n \log n)$ time and message complexity, respectively. Cheng et al. [5] proposed a polynomial time approximation scheme for the connected dominating set problem in unit-disk graphs. Huang et al. [13] formally analyze a popular algorithm used for clustering in ad-hoc mobile network scenarios. They show that this algorithm gives a 7-approximation for the minimum dominating set problem in unit-disk graphs while being able to adapt to the mobility of the nodes in the network.

Kuhn et. al. [19] presented a distributed algorithm that computes a constant factor approximation of a minimum dominating set in $O(\log^2 n)$ time without needing any synchronization but it requires that nodes know an estimate of the total number of nodes in the network. In [24], Parthasarathy and Gandhi also present distributed algorithms to compute a constant factor approximation to the minimum dominating set. The running time of their algorithm depends on the amount of information available to the nodes, and nodes have to know an estimate of the size of the network. Both papers extend the unit-disk model taking into account signal interference. A more realistic model taking physical carrier sensing into account was considered by Kothapalli et al. [16], but their algorithm needs $O(\log^4 n)$ time steps, w.h.p., in order to construct a constant factor approximation of a minimum dominating set.

4. COMMUNICATION MODEL

In this section, we present the communication model used in the design of our algorithm. The radio propagation and physical carrier sensing components of the model we use were first proposed by Kothapalli et al. in [16]. The interference model component is similar to the one used in [23].

4.1 Signal propagation

To model message reception, we observe that every data transmission mechanism has a minimum Signal-to-Noise-Interference Ratio (SINR) at which a data frame can still be received with a reasonably low frame error rate. The minimum SINRs for 802.11b, for example, are 10dB for 11Mbps down to 4dB for 1Mbps. These SINR values specify the *transmission range* of the data transmission mechanism, i.e. the maximum range within which data frames can still be received correctly. The transmission range, however, is highly dependent on the environment.

In order to handle non-uniform environments, we propose the following cost model. Assume we are given a set V of mobile stations, or *nodes*, that are distributed in an arbitrary way in a 2-dimensional Euclidean space. For any two nodes $v, w \in V$ let $d(v, w)$ be the Euclidean distance between v and w. Furthermore, consider any cost function c with the property that there is a fixed constant $\theta \geq 0$ so that for all $v, w \in V$,

$$c(v, w) \in [(1 + \theta)^{-1} \cdot d(v, w),\ (1 + \theta) \cdot d(v, w)] \,. \qquad (1)$$

The cost function c determines the transmission and interference behavior of nodes, and θ bounds the non-uniformity of the environment. In particular, transmission from node v to w is considered successful in our model (in absence of interference) if and only if $c(v, w) \leq r_t$, where r_t is the intended transmission range. Notice that we do not require c to be monotonic in the distance, to satisfy the triangle inequality, nor to be symmetric. This makes sure that our model even applies to highly irregular environments. Similar cost functions were also used in [16, 18, 23], for example.

Note that if c is not symmetric, then we have to rephrase the definition of a dominating set to avoid ambiguities. In that case, a dominating set is a set U of nodes so that for every node v, either $v \in U$ or v has a node $u \in U$ with $c(u, v) \leq r_t$ (i.e., v can receive a message from u).

It is worth observing that, by properly setting the constant θ, the above model can be used to represent a channel propagation model which is similar to the well-known log-normal shadowing model [27], in which the received power at a distance of d relative to the received power at a reference distance of d_0 (representing the distance at which the signal strength starts to degrade) is given in dB as

$$-10 \log(\max\{d/d_0, 1\})^\alpha + X_\sigma \qquad (2)$$

where α is the path loss coefficient and X_σ is a Gaussian random variable with zero mean and standard deviation σ (in dB) that models variability in signal loss with distance. α usually ranges from 2 (free space) to 5 (indoors), and σ from $2dB$ to $8dB$.

Note that the original log-normal shadowing model cannot be represented through the notion of link cost defined in (1), since the random component of signal propagation (variable X_σ in equation (2)) has infinite support. This implies that, in principle, it is possible to communicate to nodes which are arbitrarily distant from the transmitter (or to not being able to communicate to nodes which are arbitrarily close to the transmitter). Hence, the log-normal shadowing model cannot be represented by any notion of link cost which confines possible successful transmission to a pair of nodes whose distance is within a constant factor from the intended communication range. To circumvent this problem, we consider a bounded version of the log-normal shadowing model, in which the random component of signal propagation is represented by a random variable X'_σ with bounded support. In particular, the support of X'_σ is of the form $[-h \cdot \sigma, +h \cdot \sigma]$, where σ is the standard deviation of variable X_σ, and h is a constant. The pdf of X'_σ is obtained from the pdf of X_σ by uniformly distributing the probability density of variable X_σ falling outside the support of X'_σ into the interval $[-h \cdot \sigma, +h \cdot \sigma]$. It is easy to see that, by increasing h, we have that the pdf of X'_σ becomes arbitrarily close to the pdf of the original variable X_σ. For instance, by setting $h = 3$, we have that only 0.0027 of the probability mass of variable X_σ falls outside $[-3\sigma, 3\sigma]$, and the pdf of X'_σ is virtually indistinguishable from the pdf of X_σ.

It is easy to see that the above described bounded version of the log-normal shadowing model can be represented by the notion of link cost defined in (1) by setting $\theta = 10^{\frac{h\sigma}{10\alpha}} - 1$, where σ and α are the parameters of the propagation model. For instance, by setting $\alpha = 3$, $\sigma = 6dB$, and $h = 3$, we obtain $\theta \approx 1.5$, implying that a transmission between nodes u and v is always successful when $d(u, v) < 0.399r_t$, and that a successful transmission can only occur at distance $\leq 2.5r_t$. In summary, our cost model can be used to cover a bounded variant of the log-normal shadowing model, but in this case the cost function c would be a random variable. In order to be able to eventually arrive at a stable dominating set, we will only consider fixed cost functions c in this paper.

In the following, we assume that nodes use some fixed-rate communication mechanism and operate over a single frequency band. When a node u is using a transmission power of P (the same for every node), then the received power at node v is equal to $P_v(u) = P/\max\{c(u, v)^\alpha, 1\}$ where α is the path-loss exponent[2]. Similarly to [4, 12], we will assume throughout the paper that $\alpha > 2 + \epsilon$ for some arbitrary constant $\epsilon > 0$, which is usually the case in reality.

4.2 Interference model

In this paper, we model interference using the physical interference model introduced in [12], which accounts for the SINR at the receiver end of a link to determine whether the transmission is successful. More specifically, a message sent by node u to node v is correctly received if and only if

$$\frac{P_v(u)}{N + \sum_{w \in S} P_v(w)} \geq \beta \qquad (3)$$

where $P_x(y)$ is the received power at node x of the signal transmitted by node y, N is the background noise, S is the subset of nodes in $V \setminus \{u, v\}$ that are currently transmitting, and β is a constant that depends on the desired rate, the modulation scheme, etc.

It has been observed in the literature that, when using forward error correction mechanisms as proposed in the IEEE 802.11e MAC standard currently under development, the transition between being able to correctly receive a data frame with high probability and not being able to correctly receive a data frame with high probability is very sharp. As shown in [6], it can be less than 1 dB. Thus, the transmission range is an area with a relatively sharp border as implied by (3) that in reality, however, may have a very irregular shape due to environmental effects. These features (irregular coverage area, and sharp transition between low and high reception probability) are well captured by the link cost model defined in (1), and by the interference model defined in the next sub-section.

Combining the physical interference model with the radio propagation model introduced in the previous section, we have that a message sent by node u to node v is correctly received if and only if

$$\frac{\frac{P}{\max\{c(u,v)^\alpha, 1\}}}{N + \sum_{w \in S} \frac{P}{\max\{c(w,v)^\alpha, 1\}}} \geq \beta \ .$$

Observe that the physical interference model, contrary to the case of simpler interference models such as graph-based and protocol models, accounts also for the interference generated by nodes which are far away from the intended receiver (say, node v) of a communication. Although the contribution to the interference level at v of a single far-away transmitter can be relatively small, the overall contribution of *all* far-away transmitters can be sufficiently high to drive the SINR at v below the threshold and corrupt transmission. This is the reason why protocols based on localized interference models that simply ignore interference beyond a certain range from the transmitter are not guaranteed to work in a real scenario, where actual message reception probability is governed by SINR.

The fact that interference in the real world cannot, in principle, be confined within a bounded region from the transmitter poses a major challenge to the design of distributed protocols for multi-hop networks. In fact, *locality* i.e., the ability of designing protocols based on message exchange only between nodes which are at most a few hops away from each other in the network topology, is fundamental to ensure that the designed distributed protocol runs effectively even in large networks (e.g., sensor networks).

To get around this apparent contradiction between locality and the use of the physical interference model, we make extensive use of physical carrier sensing. As we show in this paper, by properly tuning the carrier sensing threshold it is indeed possible to design localized, fully distributed protocols which are guaranteed to function correctly (w.h.p.) under the physical interference model.

4.3 Physical carrier sensing

In this paper, we assume that nodes can perform physical carrier sensing, and that they can set the carrier sensing threshold to different values.

Physical carrier sensing is part of the 802.11 standard, and is provided by a Clear Channel Assessment (CCA) circuit. This circuit monitors the environment to determine when it is clear to transmit. The CCA functionality can be programmed to be a function of the Receive Signal Strength Indication (RSSI) and other parameters. The RSSI measurement is derived from the state of the Automatic Gain Control (AGC) circuit. Whenever the RSSI exceeds a certain threshold, a special Energy Detection (ED) bit is switched to 1, and otherwise it is set to 0. By manipulating a certain configuration register, this threshold may be set to an absolute power value of t dB, or it may be set to be t dB above the measured noise floor, where t can be set to any value in the range 0-127. The ability to manipulate the CCA rule allows the MAC layer to optimize the physical carrier sensing to its needs. Adaptive setting of the physical carrier sensing threshold has been used, for instance, in [30] to increase spatial reuse.

[2] The max operator is used to account for the fact that the log-distance path loss model holds only for distances beyond a certain close-by distance d_0 [27], w.l.o.g. assumed to be 1 in the following.

Parameter	Meaning	Variable	Meaning
θ	constant > 0 for defining link cost	T	threshold for physical carrier sensing in TWIN
α	path loss exponent ($\alpha > 2$)	D	upper bound on max twin density
N	background noise	d	upper bound on transmitters density
β	SINR thr. for correct message reception	$acc(v)$	account variable for node v ($acc(v) > 0$ iff v is active)
P	nodes transmission power	p_v	tx probability for node v (stage 2)
$r_t = \sqrt[\alpha]{\frac{P}{\beta N}}$	nodes transmission range	$\hat{p}$	max tx probability value
$r_s = \rho r_t$	range for physical carrier sensing ($0 < \rho < 1$)	γ	increase/decrease step of node tx probability

Table 1: Network model parameters and TWIN algorithm variables.

In our network model, nodes can not only send and receive messages, but also perform physical carrier sensing. Given some sensing threshold T (that can be flexibly set by a node), a node v senses a busy channel if and only if

$$N + \sum_{w \in S} P_v(w) \geq T ,$$

where S is the subset of nodes in $V - \{v\}$ that are currently transmitting when node v is sensing the channel.

4.4 Summary of network model

Summarizing, we consider a wireless network where correct message reception at the receiver end of a transmission is determined by the experienced SINR value computed according to equation (3), and radio signal propagation is expressed in terms of *i*) a fixed cost of a communication link as defined in (1), and *ii*) a signal loss exponent α. Finally, we assume that nodes can perform physical carrier sensing, and that the threshold used to sense the channel can be chosen among a (sufficiently large) set of possible values.

The main parameters of our network model, as well as variables used in the TWIN protocol, are summarized in Table 1.

5. THE TWIN PROTOCOL

We assume that all nodes transmit with some fixed, uniform transmission power P. Let r_t be the transmission range of that power, i.e., under an ideal situation (only the background noise is there), v can transmit a message to node w if and only if $c(v, w) \leq r_t$. In other words, r_t satisfies $P/r_t^\alpha \geq \beta N$.

Our dominating set protocol, called TWIN, is based on two carrier sensing thresholds:

- Threshold T_s is defined so that whenever a node w with $c(v, w) \leq r_s$ sends a message for some sensing range r_s, node v will notice a busy channel. We assume that $r_s = \rho r_t$ for some small constant $0 < \rho < 1$ that satisfies $T_s = (P/(\rho r_t)^\alpha) \geq 4N$ so that the sensing threshold is sufficiently far above the noise floor.
- Threshold $T_n(d)$ is defined so that if the density of transmitting nodes is at most some constant d, then a node will only sense a busy channel if there is a sending node within its transmission range.

The following lemma bounds $T_n(d)$ in terms of α and the maximum density d of nodes we expect.

LEMMA 5.1. *If $\alpha > 2$, then we can set $T_n(d) = N + C(\alpha, d)$ with*

$$C(\alpha, \delta) = \delta \cdot P \cdot \frac{\pi(1+\theta)^\alpha}{r_t^{\alpha-2}} \cdot \frac{2}{2^{\frac{\alpha}{2}} - 2}$$

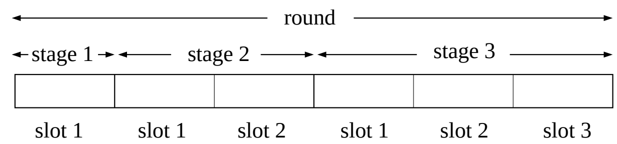

Figure 1: Round of execution of the TWIN protocol.

PROOF. To prove the lemma, we need to show that $C(\alpha, \delta)$ is an upper bound for the total signal strength caused by the nodes outside of the transmission range of u, for a maximum density value δ. This follows along the lines of [4]. □

The TWIN protocol works in *rounds* that are continuously executed and synchronized among the nodes. A node can be either *inactive* or *active*, and active nodes can be either *singles* or *twins*, as will be explained later in this section. The active nodes will eventually converge to a dominating set.

Each round of the TWIN protocol consists of three stages (see Figure 1). In stage 1, the active twins send out an ACTIVE signal with a certain probability so that inactive or active singles can learn about active twins in their vicinity. In stage 2, those nodes v that have not yet found an active twin in their vicinity probe the wireless medium and adjust their probabilities p_v so that within a certain number of rounds the sum of the probabilities within any transmission range of a node is within a constant on expectation. In stage 3, the non-twin nodes that were able to receive each other's signal in stage 2 acknowledge this to each other to be sure to form active twins. They will then announce it to the other nodes so that close-by nodes terminate the protocol.

In order to become an *active single*, each node v maintains an account $acc(v) \geq 0$. Each time $p_v = \hat{p}$, the maximum transmission probability value a node can have, it sets $acc(v) := acc(v) + 4$, and each time $p_v < \hat{p}$, it sets $acc(v) := \max\{acc(v) - 1, 0\}$. A node is an active single as long as $acc(v) > 0$.

Next, we give the details of our TWIN protocol. Initially, all nodes are inactive and $acc(v) = 0$ for every node v. The probability values p_v may be set to any value x with $0 < x \leq \hat{p}$. Each round works as follows

- **Stage 1: Announcing active twins**
  This stage consists of one time slot. In that time slot, each active twin v decides with probability $1/D$ to send out an ACTIVE signal, where the constant D is an upper bound on the maximum density of twins determined later. Each inactive or active single v that *receives* an ACTIVE signal stops executing the protocol (since it is covered) and sets $acc(v) := 0$ (i.e., it becomes inactive).
- **Stage 2: Guessing the right density**
  This stage consists of two time slots. Each inactive

or active single v still participating in the protocol chooses one of the two time slots of this stage uniformly at random, say, slot s. For slot s, v decides to send a PING signal with probability p_v. If v sends a PING signal, it *senses* the wireless channel with threshold T_s in the alternative slot, $\bar{s}$. Otherwise, it *senses* the wireless channel with threshold T_s in both slots. If it does not sense anything in either case, it sets $p_v := \min\{(1+\gamma)p_v, \hat{p}\}$, and otherwise it sets $p_v := (1+\gamma)^{-1}p_v$ for some constants $\hat{p} < 1$ and $0 < \gamma < 1$. Whereas γ may be set to any constant value, our analysis requires that $\hat{p} \leq 1/(240\pi(1+\theta)^4)$, but we did not try to optimize this bound. If $p_v = \hat{p}$, then $acc(v) := acc(v)+4$, (i.e., v becomes or remains an active single) and otherwise $acc(v) := \max\{acc(v)-1, 0\}$.

- **Stage 3: Forming new twins**
  This stage consists of three time slots. Every inactive or active single v that sent a PING signal in some slot s and *received* a PING signal in the alternative slot $\bar{s}$ does the following. It sends an ACK signal in slot s of this stage and listens to the wireless channel in slot $\bar{s}$ of this stage. If it *receives* an ACK signal in slot $\bar{s}$, it becomes an active twin.

  All nodes that just became a new active twin in this stage send a NEW signal in the last time slot. All remaining inactive or active singles *sense* the wireless channel with threshold $T_n(d)$, where d is the maximum density at which new active twins can emerge. Each node v that senses a busy channel stops executing the protocol and sets $acc(v) := 0$ (since it has an active twin within its transmission range).

Hence, altogether, each round consists of six time slots. The assumption that the rounds are synchronized among the nodes is not needed any more if six frequencies are available, one frequency for each of the six time slots of the protocol. In this case, we would only need the assumption that the drift between the local clocks of the nodes is sufficiently small for the protocol to work.

In order to avoid switching back and forth between the two sensing thresholds all the time, TWIN may just use a single threshold $T := \max\{T_s, T_n(d)\}$. This is not a problem for the protocol as long as $T_n(d) < P$ so that a single, sufficiently nearby node is able to trigger a busy channel. According to the bound in Lemma 5.1, $T_n(d) < P$ is true if r_t is sufficiently large. With a much more careful analysis that takes into account that all nodes sending NEW signals must have received ACK signals before and the convexity of the signal propagation function, it turns out that an upper bound of $2(2\pi((1+\theta)/r_t)^{\alpha-1} + 1/\beta)$ suffices for the sensing threshold in stage 3 so that a node only senses a busy channel if a NEW signal is sent by a node within its transmission range. Thus, relatively small constants r_t and β already suffice. We will defer the proof of that bound to the full version of the paper. In the following analysis, we will just argue with the T above.

5.1 Analysis of the TWIN protocol

Next, we prove the main result of this paper, i.e., we show that our protocol constructs a constant density dominating set within $O(\log(n + 1/\varphi))$ rounds, w.h.p., where φ is the lowest probability value a node v has at the beginning of the protocol. Hence, when initially setting $p_v = \hat{p}$ for all v, then the runtime bound is $O(\log n)$.

Let R_v be the current set of inactive or active singles within the transmission range of node v and $R_s(v)$ be the current set of inactive or active singles within the sensing range of node v with threshold T (i.e., whenever a node $w \in R_s(v)$ transmits a message, v will sense a busy channel with threshold T). According to the definition of T, $R_s(v) \subseteq R_v$. We need a series of lemmas to prove the theorem. The first lemma implies that after a logarithmic number of rounds a point is reached so that most of the time, $\sum_{w \in R_s(v)} p_w$ is bounded from above by a constant (Lemma 5.3). With the help of this result we can prove an upper bound on the expected number of rounds in which $\sum_{w \in R_s(v)} p_w$ is above a constant (Lemma 5.5). The two lemmas can then be used to show that most of the time the expected interference at a node caused by nodes outside of its transmission range is below $T/2 - N$ (Lemma 5.6). Hence, most of the time, $\sum_{w \in R_v} p_w = O(1)$ and the interference caused by nodes outside of R_v together with the noise is less than T (Lemma 5.7). That will allow us to prove that within a logarithmic number of communication rounds, any node v will either be an active single or have an active twin within its transmission range (Lemma 5.8). Finally, we show that within a logarithmic number of communication rounds there can only be a constant number of active singles and twins within the transmission range of v (Lemma 5.9). All of the lemmas hold w.h.p. Combining them yields Theorem 2.1.

For the analysis of our protocol we need the following general form of the Chernoff bounds [28].

LEMMA 5.2. *Consider any set of binary random variables $X_1, \ldots, X_n$, and let $X = \sum_{i=1}^n X_i$. If there are probabilities $p_1, \ldots, p_n$ with $\mathrm{E}[\prod_{i \in S} X_i] \leq \prod_{i \in S} p_i$ for every set $S \subseteq \{1, \ldots, n\}$, then it holds for $\mu = \sum_{i=1}^n p_i$ and any $\delta > 0$ that*

$$\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu \leq e^{-\frac{\delta^2\mu}{2(1+\delta/3)}}$$

If, on the other side, there are probabilities $p_1, \ldots, p_n$ with $\mathrm{E}[\prod_{i \in S} X_i] \geq \prod_{i \in S} p_i$ for every set $S \subseteq \{1, \ldots, n\}$, then it holds for $\mu = \sum_{i=1}^n p_i$ and any $0 < \delta < 1$ that

$$\Pr[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^\mu \leq e^{-\delta^2\mu/2}$$

We assume for stage 2 that, as a worst case, the set of inactive and active singles participating in stage 2 is fixed. In reality, it decreases monotonically, which is in our favor when proving upper bounds on sums of access probabilities in certain regions. In the following, the term 'node' refers to inactive or active singles still participating in the protocol.

LEMMA 5.3. *Let R be any region with the property that for any two nodes $v, w \in R$ it holds that v and w are within the sensing range of each other. Consider any time interval I in which initially $\sum_{v \in R} p_v = \phi_0$ and let X_ϕ be a random variable denoting the number of rounds in I in which $\sum_{v \in R} p_v \geq \phi$. Then it holds for any $\phi \geq 2$ and $\delta \geq 2$ that*

$$\Pr[X_\phi \geq (1+\delta)4|I|/e^{\phi/2} + \log_{1+\gamma}\lceil\phi_0/\phi\rceil] \leq (e/(1+\delta))^{-\delta 2|I|/e^{\phi/2}}$$

PROOF. For any $v \in R$ and any round $t \geq 1$ let $p_t(v)$ denote the probability p_v used by node v at the beginning of round t. Let $p_t = \sum_{v \in R} p_t(v)$. Consider some fixed round t. The probability that v does not send a PING signal in a specific time slot s in stage 2 of that round is equal to $(1/2)(1-p_t(v))+(1/2) = 1-p_t(v)/2$ (where $(1/2)(1-p_t(v))$ is the probability that v picks s but does not transmit a PING signal in s and $(1/2)$ is the probability that v does not pick s). Thus, the probability that no node is transmitting a PING signal at some time slot s is equal to

$$\prod_{v \in R}(1-p_t(v)/2) \quad \leq \quad \prod_{v \in R} e^{-p_t(v)/2} = e^{-p_t/2} \qquad (4)$$

If both slots in stage 2 are used by PING signals in R, then $p_{t+1} = (1+\gamma)^{-1}p_t$, and therefore,

$$\Pr[p_{t+1} = (1+\gamma)^{-1}p_t] \geq \Pr[\text{both slots used in stage 2}]$$

Inequality (4) implies that the probability that at least one of the slots in stage 2 is not used by a node in R is at most $2e^{-p_t/2}$. Hence, $\Pr[\text{both slots used in stage 2}] \geq 1-2e^{-p_t/2}$ which is at least $1-2e^{-\phi/2}$ if $p_t \geq \phi$. Thus, for $p_t \geq \phi$ it holds that $\Pr[p_{t+1} = (1+\gamma)^{-1}p_t] \geq 1-2e^{-\phi/2}$.

Now, consider any interval I of t rounds, numbered from 1 to t. For round $r \in \{1,\ldots,t\}$ let the binary random variable Y_r be 0 if and only if $p_r < \phi$ or it holds that $p_r \geq \phi$ and $p_{r+1} = (1+\gamma)^{-1}p_r$. Irrespective of the previous rounds it holds that

$$\begin{aligned}
\Pr[Y_r = 0] &= \Pr[p_r < \phi] + \\
&\quad \Pr[p_r \geq \phi \ \wedge \ p_{r+1} = (1+\gamma)^{-1}p_r] \\
&= \Pr[p_r < \phi] + \\
&\quad \Pr[p_r \geq \phi] \cdot \Pr[p_{r+1} = (1+\gamma)^{-1}p_r \mid p_r \geq \phi] \\
&\geq \Pr[p_r < \phi] + \Pr[p_r \geq \phi] \cdot (1-2e^{-\phi/2}) \\
&\geq 1-2e^{-\phi/2}
\end{aligned}$$

which implies that for all $r \in \{1,\ldots,t\}$ and sets $S \subseteq \{1,\ldots,t\}$ with $s < r$ for all $s \in S$, $\Pr[Y_r = 0 \mid \prod_{s \in S} Y_s = 1] \geq 1-2e^{-\phi/2}$ and therefore $\Pr[Y_r = 1 \mid \prod_{s \in S} Y_s = 1] \leq 2e^{-\phi/2}$. From this it follows that

$$\mathrm{E}[\prod_{r \in S} Y_r] \leq (2e^{-\phi/2})^{|S|} \qquad (5)$$

for any set $S \subseteq \{1,\ldots,t\}$. Let $Y = \sum_{r=1}^{t} Y_r$. From above we know that $\mathrm{E}[Y] \leq 2t/e^{\phi/2}$. Due to (5) we can apply the Chernoff bounds in Lemma 5.2 to Y, so

$$\Pr[Y \geq (1+\delta)2t/e^{\phi/2}] \quad \leq \quad (e/(1+\delta))^{-\delta 2t/e^{\phi/2}} \qquad (6)$$

for any $\delta \geq 2$. To complete the proof, we need the following claim. Recall the definition of ϕ_0 in Lemma 5.3.

CLAIM 5.4. *The number of rounds $r \in I$ with $p_r \geq \phi$ is at most $2Y + \log_{1+\gamma}\lceil \phi_0/\phi \rceil$.*

PROOF. For any interval $I' = [t_1, t_2] \subseteq I$ with $p_{t_1-1} < \phi$ and $p_r \geq \phi$ for all $r \in I'$ it must hold that $Y_{I'} = \sum_{r=t_1}^{t_2} Y_r \geq |I'|/2$ because for every $r \in I'$ with $Y_r = 0$, $p_{r+1} = (1+\gamma)^{-1}p_r$, and for any other $r \in I'$, $p_{r+1} \leq (1+\gamma)p_r$. For the initial probability p_r of I we assumed that $p_r = \phi_0$. Hence, for the first interval $I' \subseteq I$ with $p_r \geq \phi$ for every $r \in I'$, $Y_{I'}$ must be at least $|I'|/2 - \log_{1+\gamma}\lceil \phi_0/\phi \rceil$ so that indeed $p_r \geq \phi$ for every $r \in I'$. Thus, altogether it must hold for the number of rounds $r \in I$ with $p_r \geq \phi$ that these are at most $2Y + \log_{1+\gamma}\lceil \phi_0/\phi \rceil$. □

Inequality 6 and Claim 5.4 imply that

$$\begin{aligned}
&\Pr[X_\phi \geq (1+\delta)4t/e^{\phi/2} + \log_{1+\gamma}(1+\phi_0)] \\
&\quad \leq \quad \Pr[Y \geq (1+\delta)2t/e^{\phi/2}] \\
&\quad \leq \quad (e/(1+\delta))^{-\delta 2t/e^{\phi/2}}
\end{aligned}$$

for any $\delta \geq 2$, which completes the proof of Lemma 5.3. □

Lemma 5.3 allows us to bound the expected number of time steps in which $p_t \geq \phi$ for some time interval I. The bound we present is not obvious since we want to avoid the additive term of $\log_{1+\gamma}\lceil \phi_0/\phi \rceil$ in Lemma 5.3, which can be as large as $\log_{1+\gamma}(n\hat{p})$.

LEMMA 5.5. *Let R be any region with the property that for any two nodes $v, w \in R$ it holds that v and w are within the sensing range of each other. Consider any sufficiently large time interval I with $|I| = \Omega(\log_{1+\gamma} n)$ starting at a sufficiently large time step $t_0 = \Omega(\log_{1+\gamma} n)$ and let X_ϕ be a random variable denoting the number of rounds in I in which $\sum_{v \in R} p_v \geq \phi$. Then it holds for any $\phi \geq 2$ that*

$$\mathrm{E}[X_\phi] \leq 40|I|/e^{\phi/2}$$

PROOF. For any time step t let $p_t = \sum_{v \in R} p_t(v)$. Initially, $p_t \leq n$. Hence, it follows from the proof of Lemma 5.3 that for any sufficiently large interval I' of size $\Theta(\log_{1+\gamma} n)$ and any $\phi \geq 2$ there is a time step $t \in I'$ with $p_t \leq \phi$ w.h.p. Thus, if interval I starts at a sufficiently large time step $t_0 = \Omega(\log_{1+\gamma} n)$, there must be a time step $t \leq t_0$ with $t_0 - t = O(\log_{1+\gamma} n)$ and $p_t \leq \phi$ w.h.p. Let I' be the interval starting at t and containing I. Then we know from Lemma 5.3 that when defining X_ϕ with respect to I',

$$\Pr[X_\phi \geq (1+\delta)4|I'|/e^{\phi/2}] \leq (e/(1+\delta))^{-\delta 2|I'|/e^{\phi/2}}$$

for any $\delta \geq 2$. The same bound also applies to I. If $|I'| \leq (5/4)|I|$ (which is true if $|I|$ is sufficiently large, w.h.p.), then it follows that

$$\Pr[X_\phi \geq (1+\delta)5|I|/e^{\phi/2}] \leq (e/(1+\delta))^{-\delta 2|I|/e^{\phi/2}}$$

for any $\delta \geq 2$. When defining $\ell = (1+\delta)5|I|/e^{\phi/2}$ for some $\delta \geq 3$, it holds that

$$\delta 2|I|/e^{\phi/2} = \frac{2\delta}{5(1+\delta)}\ell \geq (3/10)\ell$$

and therefore for any such ℓ, $\Pr[X_\phi \geq \ell] \leq (e/4)^{(3/10)\ell} \leq 2^{-\ell/6}$. Hence, for $\ell_0 = 20|I|/e^{\phi/2}$,

$$\begin{aligned}
\mathrm{E}[X_\phi] &\leq \sum_{\ell \geq 1} \Pr[X_\phi \geq \ell] \\
&\leq \ell_0 + \sum_{\ell \geq \ell_0} \Pr[X_\phi \geq \ell] \\
&\leq \ell_0 + \sum_{\ell \geq \ell_0} 2^{-\ell/6} \leq 2\ell_0
\end{aligned}$$

□

Lemmas 5.3 and 5.5 allow us to prove the following lemma, which bounds the number of rounds in which there is too much non-local interference. Recall that the exponent for the signal degradation is α for some constant $\alpha > 2$ and that P is the transmission power of every node. Let $\bar{R}_v$ represent the complement of R_v, i.e., the area outside of the transmission range of v.

LEMMA 5.6. *Consider any time interval I and let Y be a random variable denoting the number of rounds in I in which $P\sum_{w\in\bar{R}_v} p_w/c(v,w)^\alpha \geq T/2$. For any constant $\epsilon > 0$, $\Pr[Y \geq \epsilon|I|]$ can be made polynomially small in n if $|I| = \Omega(\log_{1+\gamma} n)$ is sufficiently large.*

PROOF. For simplicity, suppose that $c(u,v) = d(u,v)$ for all pairs of nodes. Going back to the original cost model would influence the bounds below by at most a $(1+\theta)^\alpha$ factor. In the following, let $\alpha' = \alpha - 2$.

Consider $\bar{R}_v$ to be cut into rings $R_0, R_1, \ldots$ where R_i covers the area between radius $r_t + ir_s/2$ and $r_t + (i+1)r_s/2$ around v. Each ring R_i can be cut into sectors $S_{i,1}, \ldots, S_{i,k}$ with $k \leq 2\pi(r_t + (i+1)r_s/2)/(r_s/2) = 2\pi(i + 2/\rho + 1)$ for $r_s = \rho r_t$ so that every sector $S_{i,j}$ represents a region in which nodes can sense each other. Suppose that $p_t(S_{i,j}) < \phi_i$ for all i, j with $\phi_i = (i + 2/\rho + 1)^{\alpha'/2}\phi$, where the constant ϕ is specified later. Then

$$\begin{aligned}
&\sum_{w\in\bar{R}_v} p_w/d(v,w)^\alpha \\
&\leq \sum_{i\geq 0} \frac{2\pi(i+2/\rho+1)}{(r_t + i\cdot r_s/2)^{2+\alpha'}} \cdot 2(i+2/\rho+1)^{\alpha'/2}\phi \\
&= \frac{2\pi\cdot 2\phi}{(r_s/2)^{2+\alpha'}} \sum_{i\geq 0} \frac{(i+2/\rho+1)^{1+\alpha'/2}}{(2/\rho+i)^{2+\alpha'}} \\
&\leq \frac{2^{2+2\alpha'}\cdot 4\pi\phi}{r_s^{2+\alpha'}} \sum_{i\geq 0}\left(\frac{1}{i+2/\rho}\right)^{1+\alpha'/2} \\
&\leq \frac{2^{2(2+\alpha')}\cdot\pi\phi}{r_s^{2+\alpha'}} \cdot \frac{2}{\alpha'}\left(\frac{\rho}{2}\right)^{\alpha'/2}
\end{aligned}$$

Recall that $T = P/r_s^{2+\alpha'}$ for some constant P. Hence, if ρ is sufficiently small, then $P\sum_{w\in\bar{R}_v} p_w/d(v,w)^\alpha \leq T/2$.

It remains to show that most of the time this bound is true, w.h.p. Consider any time interval I starting at some sufficiently large round, and suppose that $Y \geq \epsilon|I|$. Then there are $d = \epsilon|I|$ rounds $t_1, \ldots, t_d \in I$ so that for each t_k there is an $S_{i,j}$ with $p_{t_k}(S_{i,j}) \geq \phi_i$. In order to analyze these events, let the random variable $Y_{i,j}$ to be defined as the number of times $p_{t_k}(S_{i,j}) \geq \phi_i$ in I. It follows from Lemma 5.5 that

$$\mathrm{E}[Y_{i,j}] \leq 40|I|/e^{\phi_i/2}$$

Hence,

$$\begin{aligned}
\mathrm{E}[Y] &= \sum_i\sum_j \mathrm{E}[Y_{i,j}] \\
&\leq \sum_i 2\pi(i+2/\rho+1)\cdot 40|I|/e^{\phi_i/2} \\
&\leq \epsilon|I|
\end{aligned}$$

if the constant ϕ is large enough. Each $Y_{i,j}$ can be considered as the sum of binary random variables in the flavor of the proof of Lemma 5.3 (see the definition of the Y_r's). Thus, Y can be considered as the sum of binary random variables. This sum is finite since $p(S_{i,j}) \leq n$ for every i, j, which implies that we do not have to go beyond ring i with $\phi_i > n$. Moreover, the binary random variables satisfy the condition of the first Chernoff bound in Lemma 5.2. Thus, for any $\delta > 0$,

$$\Pr[Y \geq (1+\delta)\epsilon|I|] \leq e^{-\delta^2/(2(1+\delta/2))\cdot\epsilon|I|}$$

which is polynomially small in n if $|I| = \Omega(\log_{1+\gamma} n)$ is sufficiently large. □

For a node v, a round t in I is called *good* if and only if $\sum_{w\in R_v} p_t(w) \leq g$ for some fixed constant g and the interference caused by nodes in $\bar{R}_v$ is less than $T - N$ (so that the additional noise may not trigger a busy channel).

LEMMA 5.7. *For any constant $\epsilon > 0$, at least $(1-\epsilon)|I|$ of the rounds in I are good for v, w.h.p., if g and T are sufficiently large.*

PROOF. Fix some $\epsilon > 0$. Lemma 5.3 implies that for at least $(1-\epsilon/2)|I|$ of the rounds in I, $\sum_{w\in R_v} p_t(w) \leq g$ w.h.p. if the constant g is sufficiently large. Furthermore, Lemma 5.6 implies that for at least $(1-\epsilon')|I|$ of the rounds in I, $P\sum_{w\in\bar{R}_v} p_w/c(v,w)^\alpha \leq T/2$, w.h.p., for some constant $\epsilon' > 0$ that can be arbitrarily small. Using the Chernoff bounds, only for an ϵ'-fraction of these rounds the cumulative signal strength of the nodes in $\bar{R}_v$ will exceed $T - N$ (given that $T \geq 4N$), w.h.p., so altogether there are at most $(\epsilon'(1-\epsilon') + \epsilon')|I|$ rounds in I in which the cumulative signal strength of these nodes exceeds $T - N$, w.h.p. If $\epsilon' > 0$ is sufficiently small, then $\epsilon'(1-\epsilon') + \epsilon' \leq \epsilon/2$. In this case, we would have at most an $\epsilon/2$-fraction of the rounds with $\sum_{w\in R_v} p_t(w) \leq g$ and at most an $\epsilon/2$-fraction of the rounds in which the cumulative signal strength of the nodes in $\bar{R}_v$ exceeds $T - N$, which implies the lemma. □

We are now ready to prove quick coverage by active nodes.

LEMMA 5.8. *Let I be any time interval that is starting after $\Omega(\log_{1+\gamma} n)$ rounds with $|I| = O(\log_{1+\gamma}(n+1/\varphi))$ being sufficiently large. If $\hat{p} \leq 1/(240\pi(1+\theta)^4)$ then at the end of I it holds for every node v that v is either an active single or has an active twin within its transmission range, w.h.p.*

PROOF. We distinguish between two cases for $p'_t = \sum_{w\in R_v\setminus\{v\}} p_t(w)$.

Case 1: $p'_t < 24\pi(1+\theta)^4\hat{p}$ for at least 7/8 of the rounds in I. Let t be any of these rounds and suppose that t is good. Then it follows from the proof of Lemma 5.3 that $\Pr[p_{t+1}(v) = \min\{(1+\gamma)p_t(v), \hat{p}\}]$ is at least

$$\begin{aligned}
& p_t(v)\prod_{w\in R_v\setminus\{v\}}(1 - p_t(w)/2) + \\
& \qquad (1-p_t(v))\prod_{w\in R_v\setminus\{v\}}(1-p_t(w)) \\
\geq\;& p_t(v)e^{-1.1p'_t/2} + (1-p_t(v))e^{-1.1p'_t} \\
\geq\;& e^{-1.1p'_t} \geq \frac{8}{9}
\end{aligned}$$

if $\hat{p} \leq 1/(240\pi(1+\theta)^4)$. For round t, let the binary random variable X_t be 1 if and only if $p_{t+1}(v) = \min\{(1+\gamma)p_t(v), \hat{p}\}$. Let $X = \sum_{t\in I} X_t$. If at most 1/64 of the rounds in I are bad (which is true according to Lemma 5.3 w.h.p.), it follows that $\mathrm{E}[X] \geq |I|(7/8 - 1/64)8/9 \geq |I|(3/4 + 1/72)$. Furthermore, for any good round t with $p'_t < 18\hat{p}$ and any set S of prior rounds it holds that

$$\Pr[X_t = 1 \mid \bigwedge_{r \in S} (X_r = 1)] \geq 8/9$$

Hence, we can use the Chernoff bounds to get that, if $|I| \geq c \log n$ for a sufficiently large constant c, then $X > |I|(3/4 + 1/144)$ w.h.p. If initially $p_v \geq \varphi$ and $|I| = O(\log(n + 1/\varphi))$ is large enough, then $(1+\gamma)^{|I|/144}\varphi \geq \hat{p}$, which implies together with $X > |I|(3/4 + 1/144)$ that v would have $p_t(v) = \hat{p}$ for at least a quarter of the rounds in I. When using the accounting method in the protocol (see $acc(v)$), it follows that v must be an active single at the end of I (given that no node has become an active twin within its transmission range).

Case 2: $p'_t \geq 24\pi(1+\theta)^4\hat{p}$ for at least 1/8 of the rounds in I. According to our cost model, the transmission region of v must be inside a disk of diameter $(1+\theta)r_0$ where r_0 is the ideal transmission range for $\theta = 0$. When cutting this disk into $\lceil 2\pi(1+\theta)^2\sqrt{2}\rceil$ sectors of equal angle and each sector into $\lceil(1+\theta)^2\sqrt{2}\rceil$ ring sections of equal width, then we obtain at most $8\pi(1+\theta)^4$ ring sections with diameter at most $\sqrt{2}\cdot r_0/(\sqrt{2}(1+\theta)) = r_0/(1+\theta)$. Thus, in each of these sections, any pair of nodes can communicate with each other. For each round t with $p'_t \geq 24\pi(1+\theta)^4\hat{p}$ there must be one of these sections, say S, with $\sum_{w \in S \cap R_v} p_t(w) \geq 3\hat{p}$, which implies that S can be decomposed into two groups S_1 and S_2 with $\sum_{w \in S_1 \cap R_v} p_t(w) \geq \hat{p}$ and $\sum_{w \in S_2 \cap R_v} p_t(w) \geq \hat{p}$. This implies that if t is good (so the probabilities of the nodes within the transmission range of any node in S_1 or S_2 sum up to a constant), there is a constant probability > 0 that the nodes in S will form a twin. Thus, if $|I| \geq c \log n$ for a sufficiently large constant c, then a twin will emerge within the transmission range of v, w.h.p. □

Finally, we need to prove the following lemma.

LEMMA 5.9. *At the end of time interval I, each node v can have at most a constant number of active singles and twins within its transmission range, w.h.p.*

PROOF. First, consider the twins. According to the protocol, a node v will only participate in the twin creation if it has not sensed a NEW signal in stage 3, which implies that there cannot be twins within its sensing range defined by T. Node v only becomes a twin if v sent a PING signal in stage 2 and also received one from some node w in that stage and it also receives an ACK signal in stage 3 from some node w' (that does not necessarily have to be w). If v had sent a PING signal together with a non-constant number of nodes within its transmission range, then no other node within its transmission range could have received a PING signal, which means that v would not have received an ACK signal in stage 3. Hence, only a constant number of nodes can become a twin within the transmission range of any node, and any node becoming a twin does not have a twin within its sensing range defined by T from previous rounds, so overall the twins must have a constant density.

Next, we consider active singles. Recall the cases in the proof of Lemma 5.8. Suppose that there are at least c many nodes $u \in R_v$ for which case 1 is true, which means that they would all be active singles at the end of I, w.h.p. If the constant c is sufficiently large, that would violate the case 1 assumption that $p_t < 24\pi(1+\theta)^4\hat{p}$ for at least 7/8 of the rounds in I since each of these nodes would have a probability of $\hat{p}$ for at least 1/4 of the rounds in I. Hence, there can at best be a constant number of nodes $u \in R_v$ for which case 1 applies. For the rest of them, case 2 must apply, but in this case a twin will emerge within their transmission range, w.h.p. Once a twin emerges within the transmission range of a node v, stage 1 ensures that with constant probability v will receive an ACTIVE signal from that twin in a round. Hence, these nodes will not be active singles within $O(\log n)$ rounds, w.h.p., since their accounts will be set to 0 when they terminate the protocol. □

6. TOWARDS SELF-STABILIZATION

Note that the initial probabilities of the nodes can be set in any way. From above we know that if the initial probabilities are at least φ, then the number of rounds the TWIN protocol needs is $O(\log(n+1/\varphi))$. Interestingly, also the initial $acc(v)$ values of the nodes can be set in any way since the proofs above do not make any assumptions on them besides $acc(v) \geq 0$. To see this, let us review the proof of Lemma 5.8 (the other lemmas are uncritical). In Lemma 5.8, there are two cases. If case 1 is true, node v will become an active single w.h.p. (which it will also be for very large $acc(v)$), but if case 2 is true, then v will learn about an active twin within its transmission range, w.h.p., and in this case it will set $acc(v) = 0$ and terminate the protocol, so it will be an inactive node at the end.

However, we still have two problems left. Initially, there may be a non-constant density of twins, and the nodes should not terminate the protocol at any point.

Termination is crucial for the TWIN protocol above because otherwise a non-constant density of twins may be created as the remaining nodes will continue to compete for becoming a twin. The simplest way to get around this problem is to assume that the nodes have a round threshold of $R = \Theta(\log_{1+\gamma} n)$. Whenever an inactive or active single receives an ACTIVE signal in stage 1 or senses a NEW signal in stage 3, it does not participate in stage 2 for R many time steps. This is equivalent to stopping the protocol, w.h.p., as long as no active twin will become inactive again for some reason or other events such as faults or mobility occur.

It remains to address the issue of twins with a non-constant density. An approach that may work well in practice is the following (which may be added as two additional time slots in stage 1):

Each active twin v chooses one of two time slots uniformly at random, say, slot s. For slot s, v sends out an ACTIVE signal, and for the other slot, $\bar{s}$, v *senses* the wireless medium with threshold T_a that is sufficiently high so that it will not be exceeded as long as the density of the twins does not exceed a certain constant D. If v senses a busy medium, it becomes inactive and sets $acc(v) := 0$.

Even though this approach just needs a single round to get the density distribution of the active twins down to a constant, oscillation problems can occur in pathological cases in which everywhere on expectation the cumulative signal strength is slightly below T_a. The easiest approach to avoid running into oscillation problems is that a twin sensing a busy channel sends a reset signal that is flooded to its local k-neighborhood, e.g., with the help of surviving twins, which have constant density once T_a has been applied so that the flooding effect is kept local. Every active twin receiving or sensing a reset request is becoming inactive. The

2-neighborhood of a twin should suffice to make sure that afterwards the density in its local area will not get beyond what the original TWIN protocol would create, so oscillation problems will not occur in that area any more and the dominating set will eventually stabilize.

7. CONCLUSIONS

In this paper, we have introduced a fully distributed algorithm for constructing a constant density dominating set under realistic interference and radio propagation models. We believe the ideas and techniques presented in this paper might prove very useful in the design of other distributed protocols for wireless ad-hoc networks based on realistic models, which is left for future work.

8. REFERENCES

[1] `http://standards.ieee.org/getieee802/download/802.11a-1999.pdf`

[2] M. Alicherry, R. Bathia, and L. Li, "Joint Channel Assignment and Routing for Throughput Optimization in Multi-Radio Wireless Mesh Networks", *Proceedings of ACM International Conference on Mobile Computing and Networking* (MobiCom), pp. 58–72, 2005.

[3] K. Alzoubi, P.J. Wan, O. Frieder, "New Distributed Algorithm for Connected Dominating Set in Wireless Ad Hoc Networks", *Proc. HICSS-35*, 2002.

[4] G. Brar, D. Blough, and P. Santi, "Computationally Efficient Scheduling with the Physical Interference Model for Throughput Improvement in Wireless Mesh Networks," *Proc. of ACM Mobicom*, pp. 2–13, 2006.

[5] X. Cheng, D. Huang, W. Li, W. Wu, D.Z. Du, "A Polynomial-Time Approximation Scheme for the Minimum-Connected Dominating Set in Ad Hoc Wireless Networks", *: Networks*, vol. 42, 2003.

[6] S. Choi, "IEEE 802.11e MAC-level FEC Performance Evaluation and Enhancement", *Proc. IEEE Globecom*, 2002.

[7] B.N. Clark, C.J. Colbourn, D.S. Johnson, "Unit Disk Graphs", *Discrete Mathematics*, vol. 86, pp. 165–177, 1990.

[8] D. Dubhashi, A. Mei, A. Panconesi, J. Radhakrishnan, A. Srinivasan, "Fast Distributed Algorithms for (Weakly) Connected Dominating Sets and Linear-Size Skeletons", *Proc. of ACM-SIAM SODA*, pp. 717–724, 2003.

[9] O. Goussevskaia, Y.V. Oswald, R. Wattenhofer, "Complexity in Geometric SINR", *Proc. of ACM MobiHoc*, pp. 100–109, 2007.

[10] J. Gronkvist and A. Hansson, "Comparison Between Graph-Based and Interference-Based STDMA Scheduling," *Proc. of ACM MobiHoc*, pp. 255–258, 2001.

[11] J. Gronkvist, "Traffic Controlled Spatial Reuse TDMA in Multi-hop Radio Networks," *Proc. of International Symposium on Personal, Indoor, and Mobile Radio Communications,* pp. 1203–1207, 1998.

[12] P. Gupta, P.R. Kumar, "The Capacity of Wireless Networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, 2000.

[13] H. Huang, A.W. Richa, M. Segal, "Approximation Algorithms for the Mobile Piercing Set Problem with Applications to Clustering in Ad Hoc Networks", *ACM/Baltzer MONET*, pp. 141–149, April 2004.

[14] K. Jain, J. Padhye, V. Padmanabhan, and L. Qiu, "Impact of Interference on Multi-Hop Wireless Network Performance", *Proc. of ACM MobiHoc*, pp. 66–80, 2003.

[15] T. Jurdzinski, G. Stachoviak, "Probabilistic Algorithms for the Wakeup Problem in Single-Hop Radio Networks", *Proc. Int. Symposium on Algorithms and Computation*, pp. 535–549, 2002.

[16] K. Kothapalli, C. Scheideler, M. Onus, A. Richa, "Constant Density Spanners for Wireless Ad Hoc Networks", *Proc. of ACM SPAA*, pp. 116–125, 2005.

[17] F. Kuhn, R. Wattenhofer, A. Zollinger, "Ad Hoc Networks Beyond Unit Disk Graphs", *Proc. of ACM DialM-POMC*, 2003.

[18] F. Kuhn, R. Wattenhofer, Y. Zhang, A. Zollinger, "Geometric Ad Hoc Routing: Of Theory and Practice", *Proc. of ACM PODC*, pp. 25–32, 2003.

[19] F. Kuhn, T. Moscibroda, R. Wattenhofer, "Radio Network Clustering from Scratch", *Proc. of ESA*, 2004.

[20] F. Kuhn, R. Wattenhofer, "Constant-Time Distributed Dominating Set Approximation", *Proc. of ACM PODC*, 2003.

[21] P. Kyasanur and N. Vaidya, "Routing and Interface Assignment in Multi-Channel Multi-Interface Wireless Networks", *Proc. of IEEE WCNC*, 2005.

[22] M. Luby, "A Simple Parallel Algorithm for the Maximal Independent Set Problem", *Proc. of ACM STOC*, pp. 1–10, 1985.

[23] T. Moscibroda, R. Wattenhofer, A. Zollinger, "Topology Control Meets SINR: The Scheduling Complexity of Arbitrary Topologies", *Proc. of ACM MobiHoc*, pp. 310–321, 2006.

[24] S. Parthasarathy, R. Gandhi, "Distributed Algorithms for Coloring and Domination in Wireless Ad Hoc Networks", *Proc. of FSTTCS*, 2004.

[25] B. Raman, "Channel Allocation in 802.11-based Mesh Networks," *Proc. of IEEE Infocom*, 2006.

[26] A. Raniwala and T. Chiueh, "Architecture and Algorithms for an IEEE 802.11-Based Multi-Channel Wireless Mesh Networks", *Proc. of IEEE Infocom*, pp. 2223–2234, 2005.

[27] T.S. Rappaport, *Wireless Communications: Principles and Practice, 2nd edition*, Prentice Hall, Upper Saddle River, NJ, US, 2002.

[28] C. Scheideler. *Probabilistic Methods for Coordination Problems.* HNI-Verlagsschriftenreihe 78, University of Paderborn, 2000. See also `www14.in.tum.de/personen/scheideler`.

[29] G. Sharma, R.R. Mazumdar, and N.B.Shroff, "On the Complexity of Scheduling in Wireless Networks", *Proc. of ACM Mobicom*, pp. 227–238, 2006.

[30] X. Yang, N.H. Vaidya, "Spatial Backoff Contention Resolution for Wireless Networks", *Proc. of IEEE WiMesh*, 2006.