

# The $k$ -Neighbors Approach to Interference Bounded and Symmetric Topology Control in Ad Hoc Networks\*

Douglas M. Blough<sup>†</sup>    Mauro Leoncini<sup>‡</sup>    Giovanni Resta<sup>§</sup>  
Paolo Santi<sup>§</sup>

## Abstract

Topology control, wherein nodes adjust their transmission ranges to conserve energy and reduce interference, is an important feature in wireless ad hoc networks. Contrary to most of the literature on topology control which focuses on reducing energy consumption, in this paper we tackle the topology control problem with the goal of limiting interference as much as possible, while keeping the communication graph connected with high probability. Our approach is based on the principle of maintaining the number of physical neighbors of every node equal to or slightly below a specific value  $k$ . As we will discuss in this paper, having a non-trivially bounded physical node degree allows a network topology with bounded interference to be generated. The proposed approach enforces symmetry on the resulting communication graph, thereby easing the operation of higher layer protocols. To evaluate the performance of our approach, we estimate the value of  $k$  that guarantees connectivity of the communication graph with high probability both theoretically and through simulation. We then define  $k$ -NEIGH, a fully distributed, asynchronous, and localized protocol that uses distance estimation.  $k$ -NEIGH guarantees logarithmically bounded physical degree at every node, is the most efficient known protocol (requiring  $2n$  messages in total, where

---

\*Parts of this paper appeared in Proc. ACM MobiHoc 03 [5].

<sup>†</sup>Georgia Inst. of Technology, Atlanta GA – USA. e-mail: [doug.blough@ece.gatech.edu](mailto:doug.blough@ece.gatech.edu)

<sup>‡</sup>Università di Modena e Reggio Emilia and IIT-CNR - ITALY. e-mail: [leoncini@unimore.it](mailto:leoncini@unimore.it)

<sup>§</sup>IIT – CNR, Pisa – ITALY. e-mail: [{giovanni.resta, paolo.santi}@iit.cnr.it](mailto:{giovanni.resta, paolo.santi}@iit.cnr.it)

$n$  is the number of nodes in the network), and relies on simpler assumptions than existing protocols. Furthermore, we verify through simulation that the network topologies produced by  $k$ -NEIGH show good performance in terms of node energy consumption and expected interference.

**Index terms:** wireless ad hoc networks, topology control, spatial reuse, energy consumption, connectivity.

## 1 Introduction

Topology control (TC for short) has been recently proposed as a technique to increase network capacity and to reduce energy consumption in ad hoc networks. The goal of a TC protocol is to reduce the transmission power level used by network nodes, with the constraint of preserving some fundamental properties of the communication graph (typically, connectivity). Decreasing the nodes' transmission power with respect to the maximum level potentially has two positive effects: (i) reducing the nodes' energy consumption, and (ii) increasing the spatial reuse, with a positive overall effect on network capacity [13]. Due to the limited availability of both energy and capacity in ad hoc networks, topology control is considered to be a fundamental building block of forthcoming wireless networks.

Although the potential advantages of applying TC techniques in ad hoc networks are twofold, the current literature on topology control (with the notable exception of [10], which we will discuss later) has focused attention solely on energy consumption, trying to minimize the “energy cost” of the generated (connected) network topology. This is the case, for instance, of the TC protocols presented in [1, 8, 14, 15, 16, 22, 24, 26, 28]. In these works, the issue of increasing spatial reuse, if considered at all, is addressed by providing upper bounds on the node degree in the final network topology. The rationale for considering node degree is that, if a node has relatively small degree, then it will experience relatively low contention when accessing the wireless channel. As a consequence of this, it is argued that spatial reuse is increased, as well as network capacity. However, the definition of node degree used in the current TC literature is the number of a node's one-hop neighbors in the final communication graph. Unfortunately, as we will discuss in this paper, this definition of node

degree (called *logical node degree* in the following) turns out to be inappropriate to measure the expected capacity increase due to the use of an optimized network topology. In Section 2, we provide examples supporting this claim, and we propose the notion of *physical node degree* (corresponding to the number of nodes within a node’s transmission range) that better characterizes the expected interference reduction in the final topology. More specifically, in Section 2, we prove that a small physical node degree in the communication graph results in a low expected interference at the node.

Motivated by this observation, we tackle the TC problem with the goal of generating a network topology in which the physical node degree is limited, so that network capacity is increased. More precisely, we study the problem of producing a network topology in which the physical node degree is non-trivially upper bounded (i.e., in a network with  $n$  nodes, we have an upper bound  $k$  on the physical node degree, where  $k = o(n)$ ). Given the observation above, we say that a network topology with this feature is *interference bounded*.

On the negative side, we show that producing an interference bounded network topology that preserves connectivity in the worst case (as is typically required in the current TC approaches [1, 8, 14, 15, 24, 26, 28]) is impossible. In other words, the goals of preserving worst-case connectivity and having a non-trivial upper bound on the physical node degree (and, thus, on the interference level) inherently conflict with each other.

In view of this result, and given that connectivity is often regarded as the most important feature of the network topology, one might wonder whether attempting to produce an interference bounded topology is worthwhile at all. Our belief is that producing an interference bounded topology and increasing the network capacity can be rewarding, at least in application scenarios where having a few disconnected nodes is not critical. Examples of applications of this type might be a wireless sensor network used for environmental monitoring, or a mobile ad hoc network in which users can tolerate short off-service intervals.

On the positive side, our last claim is strengthened by the theoretical and simulation-based results presented here which show that, if we exclude pathological node placements<sup>1</sup>, almost full connectivity and bounded interference can be achieved at the same time. More

---

<sup>1</sup>We recall that the observation that connectivity and bounded interference are conflicting goals is based on a worst-case analysis.

precisely, we prove that under the assumptions of uniformly distributed nodes and physical node degree upper bounded at each node by  $k = \Theta(\log n) = o(n)$ , network connectivity can be achieved with high probability (w.h.p.)<sup>2</sup>. For clarity of presentation, in the remainder of this paper we use “w.h.p.” to mean “w.h.p., under the assumption that the network nodes are distributed uniformly at random in a square area”.

In summary, our goal is to design a TC protocol that builds an interference bounded topology which is connected w.h.p. (provided the maximum power topology – i.e. the graph obtained when all the nodes transmit at maximum power – is connected). In our design, we must also fulfill the following requirements, which are fundamental if the protocol has to be implemented in practice: (i) the protocol must be asynchronous and fully distributed, and it must rely only on local information; (ii) the protocol should exchange as few messages as possible to build the communication topology; and (iii) the final topology should contain a backbone of bi-directional links. This latter property is important to favor the integration of TC with existing MAC and routing protocols [5].

To meet our design goals, we consider a topology control approach based on the generation of a symmetric subgraph of the  $k$ -neighbors graph, where  $k = o(n)$ , and  $n$  is the number of network nodes. The  $k$ -neighbors graph, i.e. the graph in which every node is connected to its  $k$ -closest neighbors, can be computed in a fully distributed and localized way, and has a non-trivial upper bound of  $k = o(n)$  on the physical node degree (i.e., it is interference bounded). Since for the reasons discussed above this communication graph may be disconnected in the worst case, we analyze its connectivity in a probabilistic setting, and we show that, assuming a uniform, random node spatial distribution, the probability of obtaining a disconnected communication graph can be made arbitrarily low.

We also present a specific protocol, called  $k$ -NEIGH, that is based on this approach and generates the desired topology in a fully distributed, asynchronous, and localized way. Our  $k$ -NEIGH protocol relies on distance estimation, a technique that can be implemented at a reasonable cost in many realistic scenarios [5]. We prove that the overall number of messages exchanged by  $k$ -NEIGH is exactly  $2n$ , and that its execution time is strictly bounded. Simulation results show that our protocol reduces energy consumption and the

---

<sup>2</sup>A certain random event  $E_n$  is said to occur with high probability if  $\lim_{n \rightarrow \infty} \text{Prob}(E_n) = 1$ .

average physical node degree considerably with respect to the case where no topology control is used, and that it compares favorably with the highly-regarded CBTC protocol of [15, 28] (which, however, guarantees worst-case connectivity).

The rest of this paper is organized as follows. In Section 2, we motivate our work, discussing the difference between logical and physical degree of a node in the communication graph. In Section 4 we give some preliminary definitions, and in Section 5 we characterize the minimum number of neighbors needed to generate a connected communication graph. In Section 6 we introduce the  $k$ -NEIGH protocol, which is a distance-estimation based implementation of the approach to topology control described in Section 5. In Section 7 we evaluate the performance of  $k$ -NEIGH through simulation. Section 8 concludes.

## 2 Motivation: logical and physical node degree

As mentioned in Section 1, the starting point of our work is the observation that the logical node degree is inappropriate to model the expected interference observed in the network.

We recall that the logical node degree is defined as the number of one-hop neighbors in the final communication topology. In most of the literature on TC, it is argued that this parameter is a measure of the expected contention at the MAC layer. This is not true, however, because the contention depends on the number of nodes in the transmission range of a given node, where the transmission range is determined by the transmission power level as set at the end of the TC protocol’s execution. We refer to the number of nodes within transmission range of a given node as the node’s *physical degree*.<sup>3</sup>

To better clarify the difference between logical and physical node degrees, consider the example depicted in Figure 1. Node  $u$  has three neighbors in the communication topology (nodes  $v, t$  and  $z$ ), so its logical degree is three. Note that some energy-inefficient links are removed in the communication topology (e.g. the link  $(u, w)$ ), so many nodes (such as node  $w$ ) do not contribute to  $u$ ’s logical degree. However, when  $u$  transmits to its farthest neighbor

---

<sup>3</sup>Although not necessary to the soundness of our definition, in this paper we assume that the path loss model is log-distance, with distance power-gradient equal to  $\alpha \geq 2$ . This ensures that, for a given power level  $p$ , all the nodes within a certain transmission range can correctly receive the message.

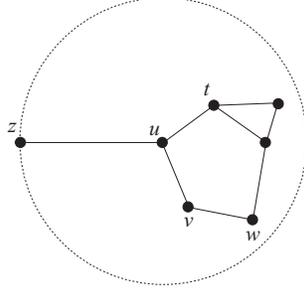


Figure 1: Difference between logical and physical node degree: node  $u$  has logical node degree equal to 3, but its physical node degree is 6.

(node  $z$ ), it interferes with *all* the nodes within its transmission range (including also node  $w$ ). For this reason, node  $w$  is accounted for when calculating  $u$ 's physical node degree. Referring to the example of Figure 1, the physical degree of  $u$  is 6. It is not difficult to build examples in which the difference between logical and physical node degree is arbitrarily large.

The following proposition outlines a fundamental tradeoff between ensuring worst-case connectivity and having a non-trivial upper bound on the physical node degree.

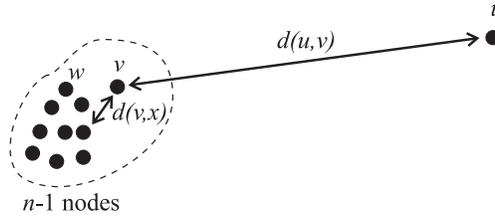


Figure 2: Example of node placement generating a connected topology where at least one node has physical degree equal to  $n - 1$ .

**Proposition 1.** *Given a set  $N$  of nodes, with  $|N| = n$ , let  $G = (N, E)$  be the communication graph obtained when all the nodes transmit at maximum power, and assume  $G$  is connected. Let  $P$  be an arbitrary topology control protocol that preserves worst-case connectivity (i.e., that generates a connected topology whenever  $G$  is connected), and let  $G_P = (N, E_P)$  be the topology generated by  $P$ . There exist node placements such that the maximum physical node degree in  $G_P$  is  $n - 1$ .*

An example of node placement generating a connected topology where at least one node has physical degree equal to  $n - 1$  is reported in Figure 2.

Informally speaking, Proposition 1 states that if we want to ensure network connectivity in the worst case, we must admit the possibility of having a “high interference node”, i.e. a node whose communications affect *all* the remaining nodes in the network.

The physical node degree as defined here has a direct relationship with the interference measure defined in [10], which is, to the best of our knowledge, the only paper tackling the TC problem with the goal of reducing interference. In [10], Burkhart, et al., introduce the notion of *interference of a graph*, which is defined as follows:

**Definition 1** (Coverage and Interference). *Let  $G = (N, E)$  be the communication graph. Given any  $(u, v) \in E$ , the coverage of edge  $e = (u, v)$ , denoted  $Cov(e)$ , is defined as the number of nodes covered by the disks induced by  $u$  and  $v$ . Formally,*

$$Cov(e) = |\{w \in N : w \text{ is contained in } D(u, d(u, v))\} \cup \{w \in N : w \text{ is contained in } D(v, d(u, v))\}| ,$$

where  $D(x, d(x, y))$  denotes the disk of radius  $d(x, y)$  centered at  $x$ .

The interference of graph  $G$  is the maximum coverage of its edges. Formally,

$$I(G) = \max_{e \in E} Cov(e) .$$

In [10], Burkhart, et al., revisit the TC problem in light of this definition of interference, and define protocols for removing high-interference edges while maintaining network connectivity. However, no explicit upper bounds on the interference of the graphs generated by these protocols are given. Actually, since the protocols preserve worst-case connectivity, the only possible bound on  $I(G)$  is  $O(n)$  (see Proposition 1, and Theorem 1 below).

The following theorem establishes a relation between the physical node degree and the interference of the communication graph.

**Theorem 1.** *Let  $G = (N, E)$  be the communication graph, and assume that the physical degree of nodes in  $G$  is at most  $k$ , for some  $k < n$ . Then, the interference of graph  $G$  is at most  $2k$ . Moreover, if the interference of graph  $G$  is at most  $k$ , then the maximum physical node degree is at most  $k - 1$ .*

*Proof.* Assume that the physical degree of nodes in  $G$  is upper bounded by  $k < n$ . Given any edge  $e = (u, v)$  in  $G$ , denote with  $S_u$  the set of nodes in  $D(u, d(u, v))$ , and with  $S_v$  the

set of nodes in  $D(v, d(u, v))$ . Since the physical degree of both  $u$  and  $v$  is at most  $k$ , it follows that  $|S_u| \leq k + 1$  (since  $S_u$  includes node  $u$ ), and that  $|S_v| \leq k + 1$  (since  $S_v$  includes node  $v$ ). By observing that at least nodes  $u$  and  $v$  are included in  $S_u \cap S_v$ , we can conclude that  $Cov(e) = |S_u \cup S_v| \leq 2k$ . Since the bound on the coverage holds for any edge  $e$  in  $G$ , the first part of the theorem follows.

Assume now that  $I(G) \leq k$ . This implies that, for any edge  $e$  in  $G$ ,  $Cov(e) \leq k$ . In turn, this implies that the physical node degree of any node  $u$  is at most  $k - 1$  (since node  $u$  itself is not accounted for in the definition of physical degree).  $\square$

Theorem 1 exposes the relationships between physical node degree and expected interference: the smaller is the physical degree, the less interference is experienced by nodes. Motivated by Theorem 1, we give the following definition of an interference bounded graph.

**Definition 2** (Interference bounded graph). *Let  $G = (N, E)$  be a communication graph with  $n$  nodes. We say that  $G$  is interference bounded if the physical node degree of nodes in  $G$  is upper bounded by  $k$ , for some  $k = k(n) = o(n)$ .*

A clarification about the notion of bounded interference used in this paper is in order. In some settings, the term “bounded” refers to a quantity which is upper bounded by a *constant* as  $n$  grows to infinity. In the definition above, we use the term “bounded” in a weaker sense, i.e. referring to a quantity which grows to infinity slower than  $n$ . Our choice of using a weaker notion of boundedness is motivated by the fact that several theoretical results have shown that if the physical node degree (i.e., interference) is upper bounded by a constant, then the resulting communication graph is disconnected w.h.p. (see, for instance, [12, 25, 29]). On the other hand, in the remainder of this paper we show that there exist communication graphs that enjoy the following properties: (i) they are connected w.h.p., and (ii) they have interference upper bounded by  $\Theta(\log n) = o(n)$ . Note also that, for practical purposes,  $\Theta(\log n)$  can be considered as a bounded quantity in the stricter sense: for instance, when  $n = 10^6$ , the logarithmic bound is less than 14 (assuming the natural logarithm).

Motivated by the above discussion, we want to design a TC protocol that generates an interference bounded topology, thus increasing network capacity. At the same time,

we want to preserve connectivity as much as possible. By Proposition 1, we know that preserving worst-case connectivity in this setting is impossible. For this reason, we weaken the connectivity constraint, requiring only that the generated topology be connected w.h.p.

Our approach to the TC problem is based on the simple idea of connecting each node to its  $k$  closest neighbors, where  $k$  is a properly tuned parameter that guarantees connectivity w.h.p. A remark on this simple idea is in order. In this paper, we formally characterize the optimal (i.e. minimum) value of  $k$ . This is done by extending the theoretical results presented in [29] to the symmetric sub-graph of the  $k$ -neighbors graph, and considering the case of deployment regions with side of arbitrary length. The selected value of  $k$  provides an upper bound on the physical degree, thereby achieving the desired property of bounded interference. Our approach also has the desirable side effect of producing symmetry on the resulting communication graph. The  $k$ NEIGH protocol that we present in Section 6 is based on this approach and is extremely efficient, exchanging only  $2n$  messages total for an  $n$ -node network. It relies only on the existence of a distance estimation mechanism, which is achievable in many settings [5].

A possible objection to the approach to topology control presented in this paper is that connecting a node  $x$  to its  $k$  closest neighbors (and only to those neighbors) might not be possible. This occurs, for instance, when  $x$  has  $k - 1$  nodes at distance less than  $d$  and two or more nodes at distance exactly  $d$ . In this case, depending on the nodes' transmission ranges, the physical degree of  $x$  would jump from less than  $k$  to strictly more than  $k$ . Generalizing this, one could conclude that no non-trivial upper bound can be devised in the worst-case. Yet, from the theoretical perspective, our approach remains valid under very reasonable assumptions. For instance, when nodes are uniformly distributed, the cumulative probability that two nodes are at *exactly* the same distance from a third node is zero. The same holds true if nodes are distributed according to other non-degenerate probability distributions (e.g. Normal, Poisson, etc.). From a more practical standpoint, if the transmission power levels are discretized, it is obviously possible to experience jumps (like the ones depicted above) in the function describing the number of neighbors. However, it is also true that node distributions such that, when increasing one level in transmission power, the physical degree jumps from  $O(1)$  or  $O(\log n)$  to  $\Theta(n)$  (i.e., from few neighbors to very many neighbors) can be regarded

as pathological. In [6] we study a version of KNEIGH tailored for the discrete power levels scenario; in our simulation results (using uniform node distributions) such pathological cases never occurred.

### 3 Related Work

The idea of maintaining a 1-hop neighborhood of a certain size has been used in the MobileGrid protocol of [17] and in the LINT protocol of [22]. Both protocols try to keep the number of neighbors of a node between low and high thresholds centered around an optimal value. When the actual number of neighbors is below (above) the low (high) threshold, the transmission range is increased (decreased), until the number of neighbors is in the proper range. However, for both protocols, no characterization of the optimal value of the number  $k$  of neighbors is given, and, consequently, no evaluation of the connectivity of the resulting communication graph is provided. Another problem of the MobileGrid and LINT protocols is that they estimate the number of neighbors by simply overhearing control and data messages at different layers. This approach has the advantage of requiring no overhead, but the accuracy of the resulting neighborhood estimate heavily depends on the traffic present in the network. In the extreme case, a node which remains silent is not detected by any of its actual neighbors. Finally, MobileGrid and LINT do not necessarily produce symmetric topologies, i.e. they can produce unidirectional links, which can present problems for other protocols in the network.

The most widely-studied TC protocol in the literature is the elegant CBTC (ConeBased-TopologyControl) protocol introduced in [28] (and further analyzed in [15]). The basic idea in CBTC is that a node  $u$  transmits with the minimum power  $p_{u,\rho}$  such that there is at least one neighbor in every cone of angle  $\rho$  centered at  $u$ . The obtained communication graph is made symmetric by adding the reverse edge to every asymmetric link. The authors show that setting  $\rho \leq 2\pi/3$  is a sufficient condition to ensure connectivity. A set of optimizations aimed at pruning energy-inefficient edges without impairing connectivity (and symmetry) is also presented. Further, the authors prove that if  $\rho \leq \pi/2$ , every node in the final communication graph has logical degree at most 6.

Compared with CBTC, our  $k$ -NEIGH protocol (which implements the  $k$  neighbors approach outlined above) relies on a weaker assumption, i.e., distance estimation vs. directional information. With respect to CBTC,  $k$ -NEIGH requires the tuning of a parameter, the preferred number of neighbors  $k$ . However, setting the proper value of  $k$  is an easy task. In fact, in Section 5 we show that the choice of  $k$ , under reasonable assumptions, is influenced only by the number  $n$  of nodes in the network, and it does not depend on the node density. In practice, setting  $k = 9$  covers a wide range of network sizes (values of  $n$  ranging from 50 to 500). Finally, we have verified that if even weaker requirements on connectivity are imposed (e.g., having at least 95% of the nodes in the largest connected component), then setting  $k = 6$  independently of  $n$  is the optimal choice.

CBTC and  $k$ -NEIGH can be seen as opposites in the following respect. CBTC preserves connectivity in the worst case but does *not* guarantee bounded physical degree (only logical degree).  $k$ -NEIGH guarantees connectivity only with a specified probability but *does* guarantee bounded physical degree. Indeed, the simulation results presented in Section 7 show that, on the average, CBTC and KNEIGH display similar performance, generating in most cases connected topologies with bounded interference.

A final difference between  $k$ -NEIGH and other protocols is that, to our knowledge,  $k$ -NEIGH is the first protocol that provides strict bounds on the number of messages exchanged and on the time taken and energy expended in determining the proper transmission power settings.

## 4 Preliminaries

Let  $N$  be a set of  $n$  nodes placed in  $[0, 1]^2$  according to some distribution. A *range assignment* for  $N$  is a positive real valued function  $RA: N \rightarrow (0, r_{max}]$  that assigns to every element of  $N$  a value in  $(0, r_{max}]$ , representing its transmission range. Parameter  $r_{max}$  is called the *maximum transmission range* of the nodes in the network and depends on the features of the radio transceivers of the nodes. We assume that all the nodes are equipped with transceivers having the same features; hence, we have a single value of  $r_{max}$  for all nodes in the network.

Given  $N$  and a range assignment  $RA$ , the *communication graph* induced by  $RA$  on  $N$  is

defined as the directed graph  $G = (N, E)$ , where the directed edge  $[i, j]$  exists if and only if  $RA(i) \geq d(i, j)$ , and  $d(i, j)$  denotes the distance between nodes  $i$  and  $j$ . In this paper, we are concerned with two variants of this graph, defined as follows:

**Definition 3.** *The symmetric super-graph of  $G$  is defined as the undirected graph  $G^+$  obtained from  $G$  by adding the undirected edge  $(i, j)$  whenever edge  $[i, j]$  or  $[j, i]$  is in  $G$ . Formally,  $G^+ = (N, E^+)$ , where  $E^+ = \{(i, j) | ([i, j] \in E) \text{ or } ([j, i] \in E)\}$ .*

**Definition 4.** *The symmetric sub-graph of  $G$  is defined as the undirected graph  $G^-$  obtained from  $G$  by removing all the non-symmetric edges. Formally,  $G^- = (N, E^-)$ , where  $E^- = \{(i, j) | ([i, j] \in E) \text{ and } ([j, i] \in E)\}$ .*

The set of neighbors of a node  $i$  in the communication graph  $G$ , denoted  $N(i)$ , is defined as the set of nodes to which  $i$  is directly connected, i.e.  $N(i) = \{j | [i, j] \in E\}$ . Neighbor sets are defined similarly in graphs  $G^+$  and  $G^-$ , the only difference being that we consider undirected instead of directed edges. Note that for these graphs  $i \in N(j)$  if and only if  $j \in N(i)$ .

Given a parameter  $k$ , with  $1 \leq k < n$ , the  $k$ -neighbors graph is the communication graph  $G_k$  in which every node is directly connected to its  $k$  nearest nodes. Formally,  $G_k$  is the communication graph induced by the range assignment  $RA_k$ , where  $RA_k(i) = d(i, j)$  and  $j$  is the  $k$ -th nearest node to node  $i$ .

It is known [19] that the power  $p_i$  required by node  $i$  to correctly transmit data to node  $j$  must satisfy  $\frac{p_i}{d_{ij}^\alpha} \geq \beta$ , where  $\alpha \geq 2$  is the *distance-power gradient* and  $\beta \geq 1$  is the *transmission quality* parameter. In ideal conditions, we have  $\alpha = 2$ ; however, in general, the value of  $\alpha$  depends on environmental conditions and is in the range  $2 \leq \alpha \leq 6$ . Setting  $\beta = 1$ , we can define the *energy cost* of a range assignment  $RA$  as  $c(RA) = \sum_{i \in N} (RA(i))^\alpha$ .

Note that the energy cost as defined above refers only to the power used in the RF amplifier, and it does not account for the power consumed in the other circuitry of the wireless card, including the receiver. We are aware that more realistic energy models have been proposed for ad hoc networks (see, for instance, [11]). However, the energy cost as defined above has been widely used in performance evaluation of TC protocols. In [7], we evaluated  $k$ -NEIGH's performance with more realistic energy models (and also considering

multi-hop data traffic). In this paper, we focus primarily on node degree when evaluating and comparing protocols, and we use this simple energy cost function as a secondary measure.

Several connectivity problems on the communication graph have been studied in the literature (see [4] and references therein). In this paper, we are concerned with the following connectivity problem on the symmetric sub-graph of the  $k$ -neighbors graph. Motivations for our interest in  $G_k^-$  can be found in Section 5.

**Definition 5** ( $k$ -neighbors Range Assignment problem, KNRA). *Let  $N$  be a set of points in a two-dimensional square region  $R$ . Determine the minimum value of  $k$  such that  $G_k^-$  is connected.*

The problem can be equivalently restated in terms of minimum energy cost; furthermore, the optimal solution can be easily found if node positions are known. In the next section, we analyze KNRA under the hypothesis that nodes are distributed uniformly at random in  $R$ . Our analysis will be used to provide a (probabilistic) guarantee on the connectivity of the topology generated by our  $k$ -NEIGH protocol.

## 5 The minimum number of neighbors for connectivity

Our approach to topology control consists in setting the nodes' transmission ranges in such a way that the resulting symmetric sub-graph  $G_k^-$  is connected w.h.p., using local information only. The choice of limiting our consideration to  $G_k^-$  is motivated by the following reasons:

- although implementing wireless unidirectional links is technically feasible (see [2, 20, 21, 23] for unidirectional link support at different layers), the actual advantage of using unidirectional links is questionable. For example, in [18], it is shown that the high overhead needed to handle unidirectional links in routing protocols outweighs the benefits that they provide, and better performance can be achieved by simply avoiding unidirectional links;
- a recent theoretical result [4] has shown that, starting from a strongly connected graph, obtaining a connected backbone of symmetric edges incurs no additional (asymptotic) energy cost.

In this section, we investigate the “preferred value” of  $k$ , i.e. the minimum value of the node degree  $k$  which guarantees connectivity w.h.p. of the communication graph. First,

we give a formal (asymptotic) characterization of this value, then we evaluate it through extensive simulations.

## 5.1 A formal characterization of the preferred value of $k$

While a formal analysis of the conditions on  $k$  under which  $G_k$  is strongly connected w.h.p. is not straightforward, the following recent result by Xue and Kumar [29] gives us the necessary technical machinery to work on its symmetric variants.

**Theorem 2.** *Assume that  $n$  nodes are placed uniformly at random in  $[0, 1]^2$ , and let  $G_k^+$  be the symmetric super-graph of the  $k$ -neighbors graph. There exist two constants  $c_1, c_2$ , with  $0 < c_1 < c_2$ , such that:*

$$\lim_{n \rightarrow \infty} \text{Prob}\{G_{c_1 \log n}^+ \text{ is disconnected}\} = 1, \text{ and}$$

$$\lim_{n \rightarrow \infty} \text{Prob}\{G_{c_2 \log n}^+ \text{ is connected}\} = 1.$$

The authors also provide explicit values for  $c_1$  and  $c_2$ , which are  $c_1 = 0.074$  and  $c_2 > 5.1774$ . Recently, the value of  $c_2$  has been improved to  $c_2 = \beta e$ , where  $\beta > 1$  is an arbitrary constant and  $e$  is the natural base [27].

Although the difference between the number of neighbors necessary and sufficient for connectivity is quite large, Theorem 2 is very important, since it states that  $\Theta(\log n)$  neighbors are necessary and sufficient for connectivity w.h.p.

Theorem 2 refers to the symmetric super-graph of  $G_k$ , in which a link that is physically unidirectional is considered as bidirectional. In other words, the connectivity of  $G_k^+$  is in general higher than that of  $G_k$ , since in  $G_k^+$  there are links that do not exist in the actual communication graph. In other words, there exist situations in which  $G_k$  is not strongly connected, but  $G_k^+$  is connected. As a consequence, the number of neighbors that is sufficient to obtain connectivity w.h.p. in Theorem 2 may not be sufficient in the actual communication graph. The following theorem and proof extend the result of Xue and Kumar to the symmetric sub-graph  $G_k$ .

**Theorem 3.** *Assume that  $n$  nodes are placed uniformly at random in  $[0, 1]^2$ , and let  $G_k^-$  be the symmetric sub-graph of the  $k$ -neighbors graph. There exist two constants  $c_1, c_2$ , with*

$0 < c_1 < c_2$ , such that:

$$\lim_{n \rightarrow \infty} \text{Prob}\{G_{c_1 \log n}^- \text{ is disconnected}\} = 1, \text{ and}$$

$$\lim_{n \rightarrow \infty} \text{Prob}\{G_{c_2 \log n}^- \text{ is connected}\} = 1.$$

*Proof.* The necessity part follows immediately by Theorem 2, since  $G_{c_1 \log n}^-$  is a sub-graph of  $G_{c_1 \log n}^+$ . To prove the sufficiency part, we have to show that the construction used in the proof of Theorem 2 holds for  $G_{c_2 \log n}^-$  also. The proof of Theorem 2 is based on the fact (proved in [29]) that any node in  $G_{c_2 \log n}^+$  is directly connected w.h.p. to every node that is within distance of  $(1 - \epsilon)r_n$ , where  $r_n = \sqrt{\frac{\eta \log n}{\pi n}}$ ,  $\epsilon$  is an arbitrary constant in  $(0, 1)$ , and  $\eta$  is a constant that depends on  $\epsilon$ . In words, this means that the communication graph  $G^{(1-\epsilon)r_n}$  generated by the  $(1 - \epsilon)r_n$ -homogeneous range assignment is a sub-graph of  $G_{c_2 \log n}^+$  (asymptotically, for  $n \rightarrow \infty$ ). Since  $G^{(1-\epsilon)r_n}$  is connected w.h.p. (for  $n \rightarrow \infty$ ) by Theorem 3.2 in [12], then  $G_{c_2 \log n}^+$  is also connected w.h.p.. The proof of our Theorem follows immediately by observing that, since any node is directly connected w.h.p. to every node that is within distance of  $(1 - \epsilon)r_n$ , and distance is obviously symmetric,  $G^{(1-\epsilon)r_n}$  is a sub-graph of  $G_{c_2 \log n}^-$  too.  $\square$

Since the proof of Theorem 3 is an extension of Xue and Kumar's theorem, the same values of the constants  $c_1$  and  $c_2$  can be used.

Having a connected backbone of symmetric edges, as provided by the  $G_k^-$  graph, allows us to use standard bidirectional link-based protocols in the upper layers, avoiding the expensive and technically difficult implementation of unidirectional links. Given the theoretical result of [4] and Theorem 3, this additional requirement on the communication graph will come with a limited additional energy cost. This statement is validated by the simulation results presented in the next subsection.

The analytical results of Theorems 2 and 3 hold under the assumption that the deployment region is fixed (it is the unit square), and the number of nodes grows to infinity. In other words, these results can be applied only to *dense* ad hoc networks, where the number of nodes per unit area is quite large. In the following, we show that the same result holds for *arbitrary* network densities in general. This generalization of Theorems 2 and 3 is very important, since it formally proves that, under the assumption that the nodes are distributed

uniformly at random in a square region, it is only the number  $n$  of nodes in the network, and not the area on which the network is deployed, that determines the preferred value of  $k^4$ .

**Theorem 4.** *Assume that  $n$  nodes are placed uniformly at random in  $[0, l]^2$ , for some  $l > 0$ , and let  $G_k^+$  be the symmetric super-graph of the  $k$ -neighbors graph. There exist two constants  $c_1, c_2$ , with  $0 < c_1 < c_2$ , such that:*

$$\lim_{n \rightarrow \infty} \text{Prob}\{G_{c_1 \log n}^+ \text{ is disconnected}\} = 1, \text{ and}$$

$$\lim_{n \rightarrow \infty} \text{Prob}\{G_{c_2 \log n}^+ \text{ is connected}\} = 1.$$

*The same result holds for the symmetric sub-graph  $G_k^-$  of the  $k$ -neighbors graph.*

*Proof.* The proof of this Theorem is reported in the Appendix. □

**Corollary 1.** *A number of neighbors in the order of  $\Theta(\log n)$  is necessary and sufficient for network connectivity w.h.p., independently of the node density. This is true for the  $k$ -neighbors graph, and for its symmetric super-graph and sub-graph.*

*Proof.* Asymptotically, we have three different regimes for the network density: high-density networks (i.e., the density grows to infinity as  $n \rightarrow \infty$ ); medium-density networks (i.e., the density converges to a constant  $c > 0$  as  $n \rightarrow \infty$ ); and low-density networks (i.e., the density converges to 0 as  $n \rightarrow \infty$ ). We already know that the statement of the corollary holds in case of high-density networks. So, we have to prove that the corollary holds also for medium and low densities.

Let us consider a sequence of sets of points:  $P_1, P_2, \dots$ , where  $P_1$  is composed of one point chosen uniformly at random in  $S_1$ ,  $P_2$  is composed of two points chosen independently and uniformly at random in  $S_{\sqrt{2/c}}$ , for some constant  $c > 0$ , and so on. In general, sequence  $P_n$  is composed of  $n$  points chosen independently and uniformly at random in  $S_{\sqrt{n/c}}$ . For any  $1 \leq k < n$ , by Lemma 4 we have:

$$\lim_{n \rightarrow \infty} \text{Pr}(G_k(P_n) \text{ is connected}) = \lim_{n \rightarrow \infty} \text{Pr}(G_k(f_{\sqrt{n/c}}(P_n)) \text{ is connected}). \quad (1)$$

---

<sup>4</sup>This generalized result is true as long as the maximum transmitting range of the nodes is high enough to ensure connectivity when all the nodes transmit at maximum power.

Since the asymptotic node density in this case is  $\frac{n}{(\sqrt{n/c})^2} = c > 0$ , and given equality (1), we have that the corollary holds also in the case of medium-density networks.

The proof for the case of low-density networks is similar, and is omitted for brevity.  $\square$

An interesting observation can be made by combining the results presented in Proposition 1 and in Corollary 1. Proposition 1 states that if the connectivity requirement on the network topology is strong (worst-case connectivity), the physical node degree can be as high as  $\Omega(n)$ . Corollary 1 states that, if we weaken the connectivity requirement by allowing a vanishingly small probability of disconnection, we can reduce the physical node degree to  $O(\log n)$ , i.e. by an exponential amount.

## 5.2 Simulation-based evaluation

The results of the previous subsection are asymptotic in nature, hence not very useful in practice. For this reason, we have also investigated the preferred value of  $k$  through extensive simulations.

In this section, the preferred value of  $k$  is selected as the minimum value of the node degree  $k$  that guarantees  $\Pr(G_k^- \text{ is connected})$  is above a certain target probability, which is set to 0.95.

The setting used for our experiments is the following. The  $n$  nodes, all with the same maximum transmission range  $R_n$ , are distributed uniformly at random in  $[0, 1]^2$ . The maximum transmission range  $R_n$  is chosen such that the communication graph that results when all nodes transmit at maximum power is connected. Details on how  $R_n$  has been set can be found in Section 7.

We have investigated the preferred value of  $k$  for different values of  $n$ . In the first experiment,  $n$  ranged from 10 to 100 in steps of 10. The reason for the small steps of  $n$  is that in most ad hoc network applications the number of nodes is expected to be in this range. For every value of  $n$ , and for every random node placement, we have calculated the minimum value of  $k$  such that  $G_k$  is strongly connected (denoted  $k_{asym}$ ), and the minimum value of  $k$  such that  $G_k^-$  is connected (denoted  $k_{sym}$ ), subject to the constraint that every node has maximum transmission range  $R_n$ . Given our choice for  $R_n$ , such minimum values for  $k$  always exist in practice. For each setting of  $n$ , we generated 100,000 random node

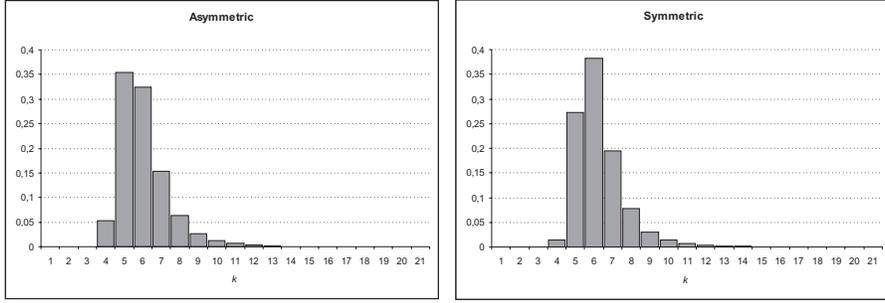


Figure 3: Empirical distribution of the minimum  $k$  for connectivity in the asymmetric (left) and symmetric (right) cases for  $n=100$ . Data are shown as frequencies.

placements, and recorded  $k_{asym}$  and  $k_{sym}$  for each of them. These data gave us the empirical probability distributions of  $k_{asym}$  and  $k_{sym}$ , which can be used to evaluate the preferred value of  $k$ . The two distributions for the case of  $n=100$  are shown in Figure 3. From the figure, it is evident that the requirement for symmetry has little influence on the minimum value of  $k$  for connectivity. This is made clearer by Figure 4, which reports the preferred value of  $k$  in the asymmetric and symmetric cases when the target probability of connectivity is set to

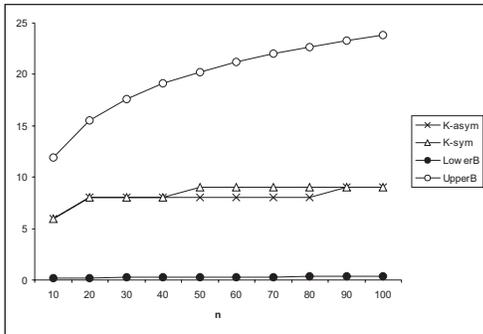


Figure 4: Preferred values of  $k$  in the asymmetric and symmetric cases ( $y$ -axis), with target probability 0.95, for different values of  $n$  ( $x$ -axis). The graphic also reports the lower and upper bounds on  $k$  derived from Theorem 3.

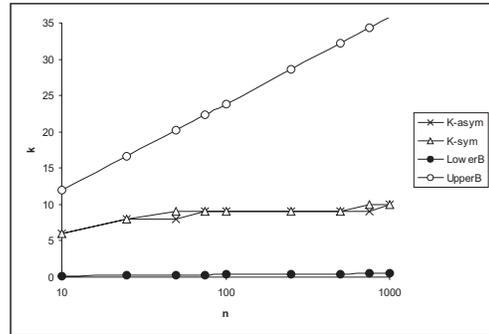


Figure 5: Preferred values of  $k$  in the asymmetric and symmetric cases ( $y$ -axis), with target probability 0.95, for different values of  $n$  ( $x$ -axis). The graphic also reports the lower and upper bounds on  $k$  derived from Theorem 3. Values on the  $x$ -axis are reported in logarithmic scale.

0.95. These values can be easily obtained by the cumulative distributions of  $k_{asym}$  and  $k_{sym}$ : the preferred value is the minimum value of  $k$  such that the cumulative frequency is above 0.95. The plots reported in Figure 4 show that the preferred value of  $k$  in the symmetric case is at most 1 greater than the value in the asymmetric case. To a certain extent, this confirms the theoretical results of Theorem 3 and of [4].

We have also evaluated how the preferred value of  $k$  varies for larger values of  $n$ . We have used the following settings for  $n$ : 10, 25, 50, 75, 100, 250, 500, 750, 1000. For every value of  $n$ , we have calculated the preferred value of  $k$  in the asymmetric and symmetric cases (with target probability 0.95), proceeding as in the previous experiment. The results of this experiment are shown in Figure 5. Again, the difference between the preferred value of  $k$  in the asymmetric and symmetric cases is at most 1, and the two values are the same for many settings of  $n$ . Interestingly, setting  $k = 9$  produces a symmetric graph which is connected with probability at least 0.95 for values of  $n$  in the range 50–500.

We have also repeated the experiments using different deployment areas, obtaining essentially the same results. This empirically confirms what was proved in the previous subsection (Theorem 4), i.e. that the preferred value of  $k$  is determined only by the number  $n$  of nodes in the network, and not by the size of the region in which the nodes are deployed.

A final investigation concerned the number of asymmetric neighbors when  $k = k_{sym}$ , i.e., in the minimal scenario for achieving connectivity in  $G_k^-$ . From our experiment, the results of which are not reported for lack of space, we observed that the average number of asymmetric links removed per node is slightly above 1.2, independently of  $n$ .

Overall, the results of this first set of simulations have shown that the requirement for symmetry has little influence on the preferred value of  $k$ , and that setting  $k = 9$  provides connectivity w.h.p. for a wide range of network sizes (from 50 to 500 nodes).

In a second set of experiments, we have weakened the connectivity requirement on the communication graph. In particular, we have re-defined the preferred value of  $k$  as the minimum value of  $k$  such that at least 95% of the nodes are in the largest connected component of  $G_k^-$ , with high probability. Here, w.h.p. means with probability at least as high as 0.95. The rationale for this investigation is that in some scenarios weaker connectivity requirements are acceptable, especially if they are counterbalanced by significant energy savings. A similar

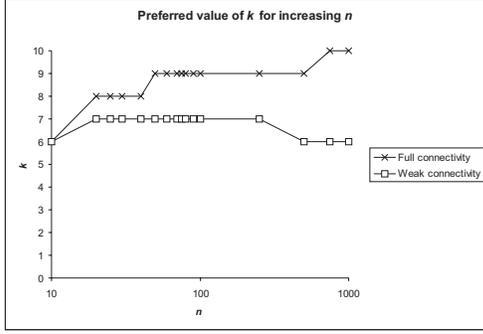


Figure 6: Preferred values of  $k$  for different values of  $n$  under strong and weak connectivity requirements. Values on the  $x$ -axis are reported in logarithmic scale.

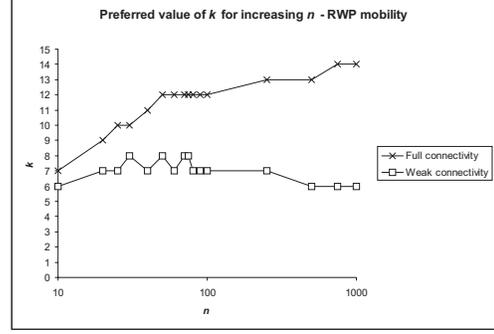


Figure 7: Preferred values of  $k$  for different values of  $n$  under strong and weak connectivity requirements in case of RWP mobility. Values on the  $x$ -axis are reported in logarithmic scale.

investigation for the critical transmission range for connectivity has been done in [25].

Figure 6 shows the preferred values of  $k$  for different values of  $n$ , under strong and weak connectivity requirements. As is seen from the figure, the type of connectivity requirement does have an influence on the preferred value of  $k$ : if we want full connectivity (all the nodes connected), then  $k$  shows an increasing behavior with  $n$  (proportional to  $\log n$ ); conversely, if we can tolerate a small percentage of disconnected nodes, then  $k$  shows a converging behavior towards the value of 6.

The results shown in Figure 6 merit discussion. The fact that weaker requirements on connectivity induce a significant reduction of the preferred value of  $k$  indicates quite clearly that the “giant component” phenomenon occurs in the  $k$ -neighbors graph. The giant component phenomenon, which is well known in the theory of geometric (and non geometric) random graphs, can be informally described as follows: assume nodes connect first to their closest neighbor, then to the second closest neighbor, and so on, until connectivity is achieved. With high probability, a large connected component (the giant component) is formed very soon in this process, and the remaining steps are needed to connect the few remaining isolated nodes to the giant component. Combining our theoretical and experimental results, we can say that if the goal is full network connectivity, the above described process stops

w.h.p. after a number of steps in the order of  $\log n$ . However, if we are satisfied with a large fraction (95%) of nodes in the largest component, we can stop the process after only 6 steps, regardless of the number  $n$  of nodes in the network.

The effect of the connectivity requirements on the preferred value of  $k$  is even more evident in case of mobile networks. It is known that some mobility models, such as the well-known random-waypoint model, generate a long-run node spatial distribution which is not uniform [3], but tends to concentrate nodes in the center of the deployment region. It is interesting to evaluate the preferred value of  $k$  in presence of RWP mobility, as the distribution resulting from this type of mobility is an instance of non-uniform, spatially concentrated node distribution, which can occur in real ad hoc networks.

To investigate the preferred value of  $k$  with RWP mobility, we have extended our simulator and performed another experiment. We have set the pause time between movements to 0, since this setting corresponds to the spatial distribution that most concentrates nodes in the center of the deployment region (see [3]). The node velocity is set to  $v = 0.01$  space units per second. We initially distributed the nodes according to the uniform distribution, and then simulated a large number (10,000) of RWP mobility steps. We then evaluated the preferred value of  $k$  on the resulting node distribution. The preferred value of  $k$  was computed both for connectivity of all nodes and for connectivity of 95% of the nodes. The results of our experiment are shown in Figure 7. As seen from the figure, the connectivity requirement plays a fundamental role: if the goal is full network connectivity,  $k$  shows an increasing behavior with  $n$ , more evident than in the stationary case: with 1,000 nodes, 14 neighbors on the average are necessary to achieve full connectivity. On the other hand, if connectivity of 95% of the nodes is sufficient, then the preferred value of  $k$  still shows a convergent behavior toward the value of 6, although with a somewhat lower convergence rate with respect to the stationary case.

Before ending this section, we remark that, not surprisingly, 6 is the magic number very famous in the networking community. As stated in [29], and confirmed by our findings, *there does not exist* a magic number of neighbors that achieves *full* connectivity in ad hoc networks. However, *if we are satisfied with connectivity of almost all nodes, 6 is still the magic number of neighbors*, both in stationary networks and in mobile networks.

## 6 The $k$ -NEIGH protocol

In this section, we describe the  $k$ -NEIGH topology control protocol – an implementation of the computation of  $G_k^-$  – and prove its correctness and complexity.

The protocol is based on the following assumptions:

1. nodes are stationary;
2. the maximum transmission power  $P$  is the same for all the nodes;
3. given  $n$ ,  $P$  is chosen in such a way that the communication graph that results when all the nodes transmit at power  $P$  is connected w.h.p.;
4. a distance estimation mechanism, possibly error prone, is available to every node;
5. the nodes initiate the  $k$ -NEIGH protocol at different times. However, the difference between node wake up times is upper bounded by a known constant  $\Delta$ .

Assumption 4 is clearly the most critical, and has been thoroughly discussed in [5], where it is also shown that  $k$ -NEIGH performance is resilient to moderate inaccuracy in distance estimation.

In the protocol specification, which is reported in Figure 8, we assume without loss of generality that the first node wakes up at time 0. At the end of the protocol execution, node  $i$  considers as neighbors (e.g., for the purpose of routing) only the nodes in the list  $L_i^S$ . Note that these are *logical* neighbors, and the set of *physical* neighbors in general is larger than  $L_i^S$  (yet bounded by  $k$ ): when  $i$  transmits at power  $P_i$ , it is possible that some node  $j \notin L_i^S$  receives the message. However, these are asymmetric neighbors, which are not considered. Also, the pruning stage reported in Figure 9 can be executed to further reduce the logical (and possibly physical) degree of some nodes, without requiring any additional messages to be exchanged.

The following results show that the  $k$ -NEIGH protocol is correct.

**Lemma 1.** *Let  $\bar{t}$  be the time necessary to transmit a message. For  $d = m\bar{t}$ , the probability that no contention will occur in the wireless channel during step 1 of the  $k$ -NEIGH protocol is  $\approx e^{-\frac{3h(h-1)}{2m}}$ , where  $h$  is the number of nodes that are contending for the channel when transmission is done at maximum power, and  $m$  is a substantially large integer (with respect to  $h$ ).*

*Proof.* In the worst case, all the nodes wake up at the same time  $\bar{\Delta} \in [0, \Delta]$ , and all the transmissions in step 1 will occur at a time taken uniformly at random in the interval  $[\Delta + \bar{\Delta}, \Delta + \bar{\Delta} + d]$ . Fix  $d = m\bar{t}$ , so that the interval  $[\Delta + \bar{\Delta}, \Delta + \bar{\Delta} + d]$  can be divided into  $m$  sub-intervals of length  $\bar{t}$  each. If node  $i$  initiates the transmission during the  $z$ -th interval (i.e., at some time in  $((z-1)\bar{t} + \Delta + \bar{\Delta}, z\bar{t} + \Delta + \bar{\Delta}]$ , for some integer  $z \in 1 \dots m$ ), we say that the  $z$ -th interval is occupied. Now, the following is clearly a sufficient condition for the occurrence of the “no contention” event: no pair of nodes occupies the same interval  $z$  and, if

---

THE  $k$ -NEIGH PROTOCOL (FOR A GENERIC NODE  $i$ ):

1. Node  $i$  wakes up at time  $t_i$ , with  $t_i \in [0, \Delta]$ . At random time  $t_i^1$  chosen uniformly in the interval  $[t_i + \Delta, t_i + \Delta + d]$  (the value of the parameter  $d$  is set in Lemma 1), node  $i$  announces its ID at maximum power;
2. For every message received from other nodes,  $i$  stores the identity and the estimated distance of the sender;
3. At time  $t_i + 2\Delta + d$ ,  $i$  orders the list of its neighbors (i.e., of the nodes from which it has received the announcement message) based on the estimated distance; let  $L_i$  be the list of the  $k$  nearest neighbors of node  $i$  (if  $i$  has less than  $k$  neighbors,  $L_i$  is the list of all its neighbors).
4. At random time  $t_i^2$  chosen in the interval  $[t_i + 2\Delta + d + \tau, t_i + 2\Delta + 2d + \tau]$  ( $\tau$  is an upper bound on the duration of step 3), node  $i$  announces its ID and the list  $L_i$  at maximum power.
5. At time  $t_i + 3\Delta + 2d + \tau$  node  $i$ , based on the lists  $L_j$  received from its neighbors, calculates the set of *symmetric neighbors* (nodes  $i$  and  $j$  are said to be symmetric neighbors if and only if  $i \in L_j$  and  $j \in L_i$ ) in  $L_i$ . Let  $L_i^S$  be the list of symmetric neighbors of node  $i$ , and let  $j$  be the farthest node in  $L_i^S$ .
6. Node  $i$  sets its transmitting power  $P_i$  to the power needed to transmit at distance  $d_{ij}^e$ , where  $d_{ij}^e$  is the estimated distance between nodes  $i$  and  $j$ .
7. (OPTIONAL PRUNING STAGE) Apply an optimization procedure to reduce the number of edges in the graph obtained so far (Figure 9).

Figure 8: The  $k$ -NEIGH topology control protocol.

an interval  $z$  is taken, then intervals  $z - 1$  and  $z + 1$  are free. Since the transmission times are independent events, we may assume that the “choices” of the transmission intervals made by nodes form a sequence of independent random variables  $Z_i$  uniformly distributed in  $[1, m]$ , with  $i = 1, \dots, h$ . A success in the  $i$ -th trial occurs when  $|Z_i - Z_j| > 1$ , for any  $j < i$ . It is easy to see that this happens with probability at least  $\frac{m-3(i-1)}{m}$ . The probability of no contention is then lower bounded by

$$\Pr\{\text{no contention}\} \geq 1 \cdot \left(1 - \frac{3}{m}\right) \cdot \dots \cdot \left(1 - \frac{3(h-1)}{m}\right)$$

Taking the logarithms and using the first term of the Taylor expansion of  $\log(1 - x)$  at  $x = 0$ , we have:

$$\begin{aligned} \log \Pr\{\text{no contention}\} &\geq \sum_{i=2}^h \log \left(1 - \frac{3(i-1)}{m}\right) = \\ &= - \sum_{i=2}^h \left[ \frac{3(i-1)}{m} + o\left(\frac{i^2}{m}\right) \right] \approx - \frac{3}{m} \sum_{i=1}^{h-1} i = - \frac{3h(h-1)}{2m}. \end{aligned}$$

The proof follows by exponentiation. □

Lemma 1 can be used to compute a lower bound to the probability that there is no contention when accessing the wireless channel. For example, if  $n = 100$  nodes are distributed uniformly at random in a square region and  $P$  is chosen in accordance with Assumption 3, the expected number of nodes within the maximum transmission range is about 33 (see Section 7 for details). Given these settings,  $d$  must be around  $16000\bar{t}$  to obtain a contention-free probability of at least 0.9. With  $\bar{t}$  in the order of, say, milliseconds,  $d$  will be in the order of tenth of seconds, which is reasonable for most topology control scenarios. Clearly, Lemma 1 provides only a crude lower bound on  $\Pr\{\text{no contention}\}$ , and smaller values of  $d$  should be usable in practice.

**Lemma 2.** *Let  $G_k^- = (N, E)$  be the undirected graph computed by steps 1-6 of the  $k$ -NEIGH protocol, and suppose  $G_k^-$  is connected. Let  $G' = (N, E')$  be the directed graph obtained as the result of the pruning stage of  $k$ -NEIGH. Then,  $G'$  is strongly connected and symmetric.*

*Proof.* We first prove that  $G'$  is strongly connected by showing that, if  $(i, j) \in E$  and  $(i, j)$  is deleted, then  $i \rightsquigarrow j$  will still hold in  $G'$ . Consider the pruning stage executed by node  $i$ . According to the protocol, node  $i$  deletes  $(i, j)$  provided that there is a neighbor  $i_1$  of  $i$  such

that  $i \rightsquigarrow_p i_1$ , for some path  $p$  in  $G_k^-$ , and moreover  $(i_1, j) \in E$ . Now, let  $p'$  denote the whole path from  $i$  to  $j$  (i.e.,  $p'$  is  $p$  plus the final edge  $(i_1, j)$ ). By the same argument above, if some edge  $(s, t)$  of  $p'$  is removed in  $G'$  as the result of the pruning stage executed by node  $s$ , then an alternative path  $p''$  exists in  $G'$  that connects  $s$  to  $t$ , and hence  $i$  to  $j$ . It is an easy consequence of the cost rule 2b of the pruning protocol that all these paths must be acyclic (for otherwise a contradiction would occur by summing the transmission powers on a circuit). Since the number of nodes is finite, the process of replacing an edge with a path must eventually stop.

As for the symmetry, it is sufficient to observe that if node  $i$  deletes  $(i, j)$ , then  $P(i, z) +$

PRUNING STAGE OF THE  $k$ -NEIGH PROTOCOL (FOR A GENERIC NODE  $i$ ):

Let  $G_k^- = (N, E)$  be the undirected graph obtained as the results of steps 1–6 of the  $k$ -NEIGH protocol, and, for any  $(i, j) \in E$ , let  $P(i, j)$  denote the transmission power sufficient for  $i$  to reach node  $j$ . This information is included in the message sent by node  $i$  during step 4.

1. Node  $i$  sorts the list  $L_i^S$  according to increasing values of  $P(i, j)$  (initially, this is equivalent to the order given by the increasing distances from  $i$ ). Let  $j_1, \dots, j_k$  be the sorted list (without loss of generality, we assume that  $L_i^S$  contains  $k$  elements; otherwise, the sorted list will be composed by  $k_1 < k$  elements).

2. For  $l = 2, \dots, k$ , do the following.

- a. Check whether  $j_l$  can be reached using a transmission power lower than  $P(i, j_l)$  by routing through some  $j_q$ ,  $q < l$ . Clearly, given the information available to node  $i$ , this is possible only if  $(j_q, j_l) \in E$ , a circumstance that is known to  $i$  from step 5 of the  $k$ -NEIGH protocol.
- b. If  $P(i, j_q) + P(j_q, j_l) \leq P(i, j_l)$ , logically delete the (outgoing) edge  $(i, j_l)$  and set  $P(i, j_l) = P(i, j_q) + P(j_q, j_l)$ . If more than one node satisfies this requirement, choose the node  $q$  such that  $P(i, j_q) + P(j_q, j_l)$  is minimum.

3. Set the transmitting power to the power needed to reach the farthest node in  $L_i^S$  which is still an immediate neighbor of node  $i$ .

Figure 9: The pruning stage.

$P(z, j) \leq P(i, j)$ , for some node  $z$ . The symmetry of the power function  $P$  implies that node  $j$  will delete  $(j, i)$  as well.  $\square$

The following Theorem summarizes the properties of the  $k$ -NEIGH protocol.

**Theorem 5.** *Assume that  $k$  is chosen in accordance with Theorem 4. Then the  $k$ -NEIGH protocol:*

- a. terminates at time at most  $4\Delta + 2d + \tau$  (where  $d$  is set in Lemma 1), i.e., by this time all the nodes have set their transmitting power correctly and terminated the protocol execution;*
- b. generates a symmetric communication graph with physical node degree at most  $k$ , which is connected w.h.p. under the hypothesis that nodes are distributed uniformly at random in  $[0, l]^2$ , where  $l$  is an arbitrary positive constant;*
- c. has communication complexity  $\Theta(n)$ . More precisely, it exchanges a total of  $2n$  messages.*

## 7 Evaluation of $k$ -NEIGH performance

In this section, we evaluate the  $k$ -NEIGH protocol through simulation. The main goal of this section is to show that, despite its extremely simple design and efficient operation, the topologies generated by  $k$ -NEIGH can provide significant improvements compared to networks without topology control, and that these improvements are, in typical cases, comparable to those provided by other protocols that rely on more accurate information about the neighborhood, such as the CBTC protocol of [15, 28] (which relies on directional information), and the LMST protocol of [16] (which relies on location information). We recall that these protocols, differently from  $k$ -NEIGH, preserve connectivity of the network topology in the worst case.

### 7.1 Simulation setup

The metrics we use for evaluation are: *energy cost* (as defined in Section 4), *physical node degree*, and *path length*. As discussed extensively earlier in the paper, physical degree is important to evaluate the expected contention at the MAC layer.

In our simulations, we have considered values of  $n$  ranging from 10 to 1,000. For each value of  $n$ , we have generated 10,000 random node placements, and executed the following

$n$	$R_n$	$n$	$R_n$
10	0.86622	75	0.37041
20	0.66420	80	0.36291
25	0.60431	90	0.34787
30	0.55589	100	0.33326
40	0.48635	250	0.23634
50	0.44526	500	0.19691
60	0.41456	750	0.17885
70	0.38336	1000	0.17274

Table 1: Values of the maximum transmission range  $R_n$  used in our simulations.

topology control algorithms:

- *MST*: although impractical (its computation requires global knowledge), the Euclidean Minimum Spanning Tree produces a range assignment that is within a factor of 2 from the optimal weakly symmetric range assignment (see [4]). We have used the MST as the “optimal” topology (from the energy consumption point of view) against which the topologies generated by the other protocols will be compared.
- *k-NEIGH*: for each setting of  $n$ , the value of  $k$  used in the protocol is the value that guarantees connectivity with probability 0.95 (as evaluated in Section 5.2).
- *CBTC*: we have simulated CBTC using two values for  $\rho$  (the maximum angular gap required):  $\rho = \frac{2}{3}\pi$  and  $\rho = \frac{\pi}{2}$ .
- *LMST*: the “local MST” protocol of [16], which computes an approximation of the MST based on local information only.
- *Homogeneous*: we have also considered the situation in which no topology control is used. In this case, the value of the transmission range is defined as the 0.95 quantile of the empirical distribution of the critical transmission range (see [25]).

The maximum transmission range used in our simulations, for every value of  $n$  considered, is shown in Table 1. As discussed in detail in [5], these values of the transmission range ensure that Assumption 3 of Section 6 is satisfied. A sample of the topologies generated by the various protocols for  $n = 100$  are shown in figures 10 and 11.

## 7.2 Energy cost

Recall that the energy cost is defined as  $c(RA) = \sum_{i \in N} (RA(i))^\alpha$ , where  $RA$  is the range assignment as defined at the end of the protocol execution. The energy cost gives a measure

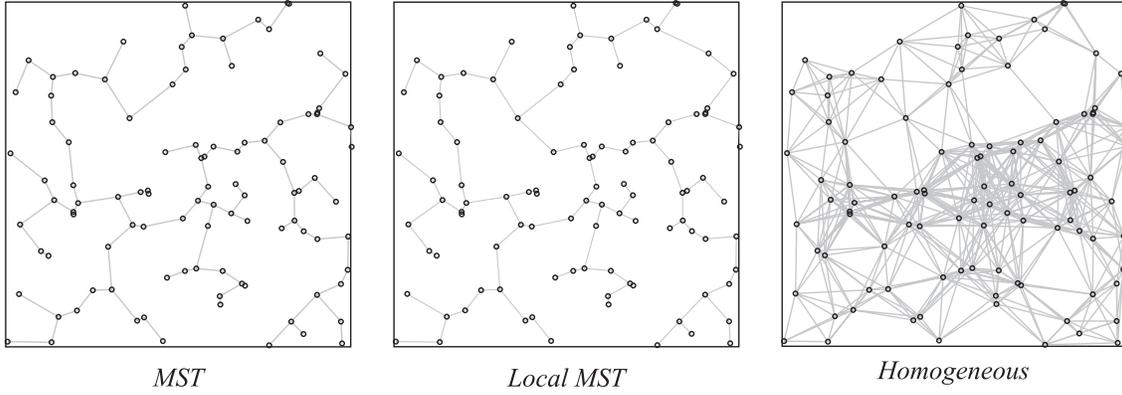


Figure 10: Sample topologies produced by the MST, LMST and homogeneous topology control protocols with  $n = 100$ .

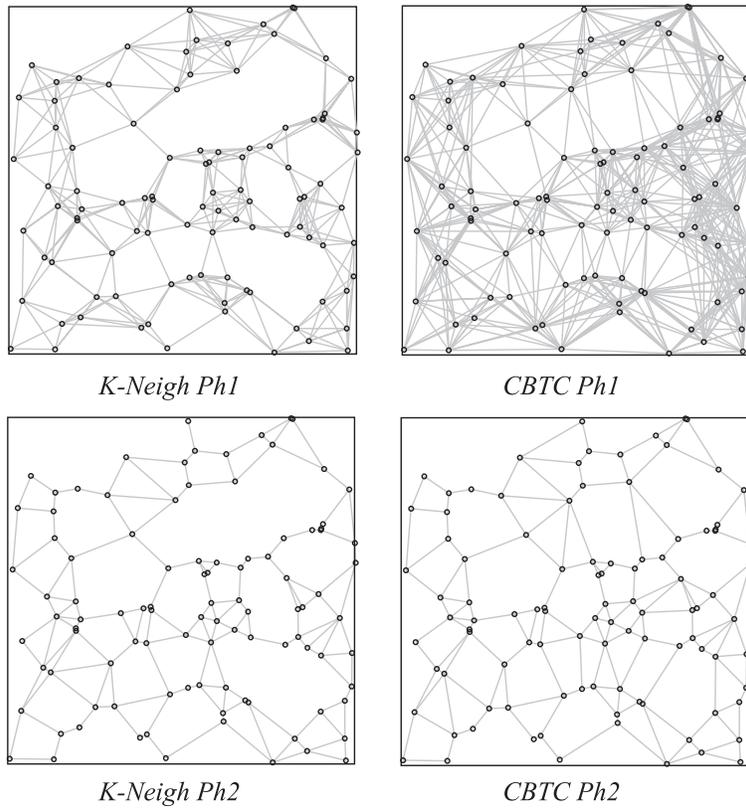


Figure 11: Sample topologies produced by the  $k$ -NEIGH and CBTC topology control protocols with  $n = 100$ . In CBTC,  $\rho$  is set to  $\frac{2}{3}\pi$ .

of the “energy efficiency” of the topology generated by a topology control algorithm.

For the  $k$ -NEIGH and CBTC protocols, we have considered both the result of Phase 1 only

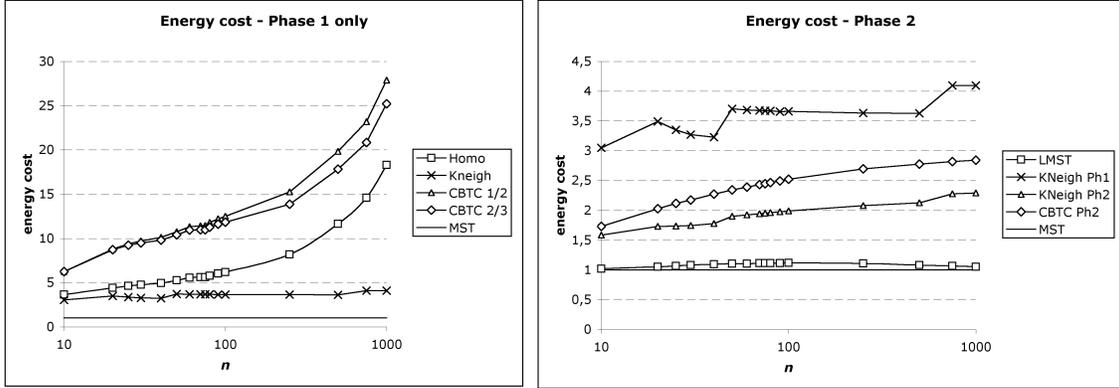


Figure 12: Energy cost of different topology control protocols. For  $k$ -NEIGH and CBTC, we have considered Phase 1 only (left), and Phases 1 and 2 implemented (right). The energy cost is normalized with respect to the cost of the MST. Values on the  $x$ -axis are reported in logarithmic scale.

(without pruning), and the result of the full protocols with the pruning phase implemented. We have considered two values for the distance-power gradient  $\alpha$ , i.e.,  $\alpha = 2$  and  $\alpha = 4$ . The value of the distance-power gradient has a strong influence on the pruning phases of  $k$ -NEIGH and CBTC, which are essentially based on triangular inequalities on the power function: the higher  $\alpha$  is, the more edges are pruned.

In Figure 12, we show the energy cost (normalized with respect to the cost of the MST) of the different protocols when  $\alpha = 2$ , for increasing values of  $n$ . To summarize the figure,  $k$ -NEIGH significantly reduces energy cost compared to networks without topology control – even without the pruning phase, it is up to 7 times better and with the pruning phase it is up to 14 times better. With the pruning phases implemented, the energy costs of  $k$ -NEIGH and CBTC are comparable, both being roughly twice that of an “optimal” protocol. It should be noted that the pruning phase of  $k$ -NEIGH can be implemented with no additional message exchanges, i.e. the total message cost of both phases combined is  $2n$  messages, which is extremely efficient. The energy cost of LMST is quite close to that of the actual MST.

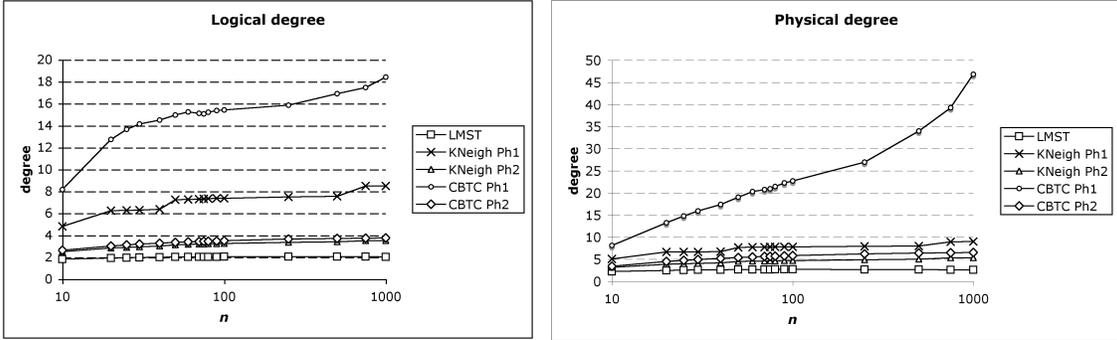


Figure 13: Average logical (left) and physical (right) degree of the topologies generated by the  $k$ -NEIGH, CBTC, and LMST protocols. Values on the  $x$ -axis are reported in logarithmic scale.

### 7.3 Physical node degree

In Figure 13, we report the average logical (left) and physical (right) node degrees of the topologies generated using  $k$ -NEIGH, CBTC, and LMST. Since LMST has been seen to perform very similarly to MST, we omit the MST results in this figure. From Figure 13, it is evident that  $k$ -NEIGH-Phase 1 outperforms CBTC-Phase 1 in terms of both logical and physical degree. Observe that in  $k$ -NEIGH we have the upper bound  $k$  on the number of *physical* neighbors of any node, which holds for Phase 1 also. The result of [28] on the logical degree (which, we recall, is 6) pertains to the topology generated by CBTC *after* pruning.

When the pruning phase (Phase 2) is considered, the performances of the two protocols in terms of average physical degree become much closer, with  $k$ -NEIGH performing only slightly better than CBTC. We recall again that, for  $k$ -NEIGH, the pruning phase comes at no additional message cost. LMST, being an approximation of the sparsest possible topology, achieves a slightly lower average physical degree compared to  $k$ -NEIGH.

### 7.4 Path length

The energy cost, as defined herein, and the average degree do not tell the whole story in terms of energy consumption and interference. In fact, while bounded degree is desirable to limit maximum interference, energy consumption and interference that consider multi-hop traffic give a better view of overall performance of a particular topology. However, such results are

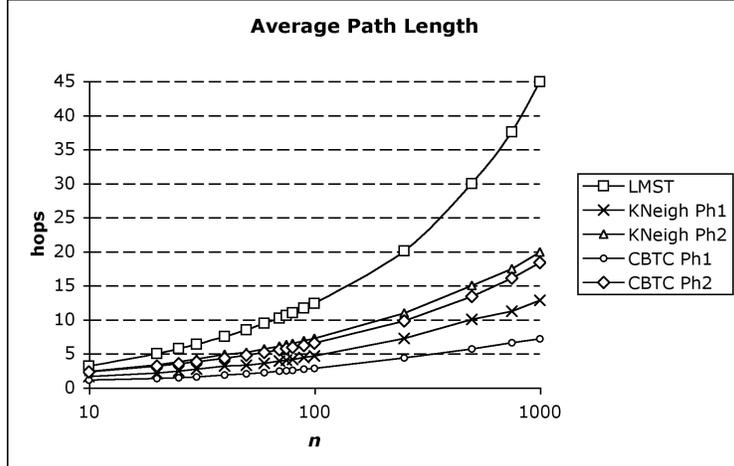


Figure 14: Average path length of the topologies generated by the  $k$ -NEIGH, CBTC, and LMST protocols. Values on the  $x$ -axis are reported in logarithmic scale.

strongly dependent on assumed traffic models, which can vary significantly from application to application. To provide a traffic-independent assessment of this aspect, we consider path length, which, when combined with physical degree and energy cost, can provide a better overall picture of topology control performance.

Figure 14 shows the average path length of  $k$ -NEIGH, CBTC, and LMST. Again, MST is omitted due to its similar performance to LMST. From Figure 14, we see that average path length increases rapidly for the very sparse topologies produced by LMST. Thus, for traffic patterns that include a modest percentage of non-local traffic, the large number of hops required for the sparsest topologies will produce significantly higher overall interference and consume greater amounts of energy than the slightly denser topologies produced by protocols such as  $k$ -NEIGH and CBTC. The results in this section demonstrate that, in absence of specific knowledge of traffic patterns, protocols such as  $k$ -NEIGH that can reduce energy cost, bound physical degree and maintain short path lengths are the best candidates for topology control.

## 8 Conclusions and future work

In this paper, we have tackled the TC problem with the goal of reducing interference between nodes while preserving network connectivity as much as possible. To this end, we have defined a quantity, the physical node degree, which is strictly related to interference, and we have designed a protocol that generates a topology with physical node degrees that are upper bounded by  $k$ . We have shown that, by setting  $k = \Theta(\log n)$  and under the assumption of randomly, uniformly distributed network nodes, the topologies generated by our protocol are connected w.h.p.

Our protocol, called  $k$ -NEIGH, is based on the simple idea of connecting each node to its  $k$  closest neighbors. We have seen that in practice  $k$ -NEIGH does not require the knowledge of the exact number  $n$  of nodes in the network to work, as  $k$  is only loosely dependent on  $n$  (e.g.,  $k = 9$  for  $n$  in the range 50-500); if a small percentage of disconnected nodes can be tolerated in the application scenario, the simulation results suggest that  $k$  can be set to 6 independently of  $n$ . Also, the maximum transmission range of nodes can be overestimated without problems, since our protocol is not influenced by the choice of a specific maximum transmission range. Whenever distance estimation is a viable choice, our protocol can be easily implemented in practice.

There are several avenues for further research on neighborhood-based topology control. Recently, we have proposed a variation of  $k$ -NEIGH which does not require distance estimation, while maintaining similar performance to the  $k$ -NEIGH protocol described herein. This variation of  $k$ -NEIGH deals also with mobile networks. We plan to investigate the performance of our protocols in the presence of multi-hop data traffic, and using a more sophisticated model for the radio signal propagation, such as that recently proposed in [9].

## 9 Acknowledgements

The work of Douglas Blough was supported in part by the National Science Foundation under Grant ECS-0225417 and INTL-0405157. The work of Paolo Santi was supported in part by CNR-NATO under grant 215.34.

## References

- [1] M. Bahramgiri, M. Hajiaghayi, V.S. Mirrokni, “Fault-tolerant ad 3-Dimensional Distributed Topology Control Algorithms in Wireless Multi-hop Networks”, *Proc. IEEE Int. Conference on Computer Communications and Networks*, pp. 392–397, 2002.
- [2] L. Bao, J.J.Garcia-Luna-Aceves, “Channel Access Scheduling in Ad Hoc Networks with Unidirectional Links”, *Proc. DIALM 01*, pp. 9–18, 2001.
- [3] C. Bettstetter, G. Resta, P. Santi, “The Node Distribution of the Random Waypoint Mobility Model for Wireless Ad Hoc Networks”, *IEEE Trans. on Mobile Computing*, Vol. 2, n. 3, pp. 257–269, July-Sept. 2003.
- [4] D.M. Blough, M. Leoncini, G. Resta, P. Santi, “On the Symmetric Range Assignment Problem in Wireless Ad Hoc Networks”, *Proc. IFIP Conf. on Theoretical Computer Science*, pp. 71–82, 2002.
- [5] D.M. Blough, M. Leoncini, G. Resta, P. Santi, “The  $k$ -Neigh Protocol for Symmetric Topology Control in Ad Hoc Networks”, *Proc. ACM MobiHoc 03*, pp. 141–152, 2003.
- [6] D.M. Blough, M. Leoncini, G. Resta, P. Santi, “K-NeighLev: a Practical Realization of Neighborhood-Based Topology Control in Ad Hoc Networks”, *Tech. Rep. IIT-TR-09/2003*, Istituto di Informatica e Telematica, Pisa - Italy, Sept. 2003.
- [7] D.M. Blough, M. Leoncini, G. Resta, P. Santi, “Comparison of Cell-Based and Topology Control-Based Energy Conservation in Wireless Ad Hoc and Sensor Networks”, *Proc. ACM MSWiM*, poster presentation, 2004.
- [8] S.A. Borbash, E.H. Jennings, “Distributed Topology Control Algorithm for Multihop Wireless Networks”, *Proc. IEEE Int. Joint Conference on Neural Networks*, pp. 355–360, 2002.
- [9] J. Bruck, M. Franceschetti, L. Schulman, “Microcellular Systems, Random Walks, and Wave Propagation”, *Proc. IEEE Symposium on Antennas and Propagation*, pp. 220–223, 2002.
- [10] M. Burkhart, P. Von Rickenbach, R. Wattenhofer, A. Zollinger, “Does Topology Control Reduce Interference?”, *Proc. ACM MobiHoc*, pp. 9–19, 2004.
- [11] L.M. Feeney, M. Nilson, “Investigating the Energy Consumption of a Wireless Network Interface in an Ad Hoc Networking Environment”, *Proc. IEEE INFOCOM 2001*, pp. 1548–1557, 2001.
- [12] P. Gupta, P.R. Kumar, “Critical Power for Asymptotic Connectivity in Wireless Networks”, *Stochastic Analysis, Control, Optimization and Applications*, Birkhauser, Boston, pp. 547–566, 1998.
- [13] P. Gupta, P.R. Kumar, “The Capacity of Wireless Networks”, *IEEE Trans. Information Theory*, Vol. 46, n. 2, pp. 388–404, 2000.
- [14] Z. Huang, C. Shen, C. Srisathapornphat, C. Jaikaeo, “Topology Control for Ad Hoc Networks with Directional Antennas”, *Proc. IEEE Int. Conference on Computer Communications and Networks*, pp. 16–21, 2002.
- [15] L. Li, J.H. Halpern, P. Bahl, Y. Wang, R. Wattenhofer, “Analysis of a Cone-Based Distributed Topology Control Algorithm for Wireless Multi-hop Networks”, *Proc. ACM PODC 2001*, pp. 264–273, 2001.
- [16] N. Li, J. Hou, L. Sha, “Design and Analysis of an MST-based Topology Control Algorithm”, *Proc. IEEE Infocom 03*, pp. 2003.
- [17] J. Liu, B. Li, “MobileGrid: Capacity-aware Topology Control in Mobile Ad Hoc Networks”, *Proc. IEEE Int. Conference on Computer Communications and Networks*, pp. 570–574, 2002.
- [18] M.K. Marina, S.R. Das, “Routing Performance in the Presence of Unidirectional Links in Multihop Wireless Networks”, *Proc. ACM MobiHoc 02*, pp. 12–23, 2002.

- [19] K. Pahlavan, A. Levesque, *Wireless Information Networks*, John Wiley and Sons, New York, 1995.
- [20] M.R. Pearlman, Z.J. Haas, B.P. Manvell, “Using Multi-Hop Acknowledgements to Discover and Reliably Communicate over Unidirectional Links in Ad Hoc Networks”, *Proc. Wireless Communications and Networking Conference (WCNC)*, pp. 532–537, 2000.
- [21] R. Prakash, “A Routing Algorithm for Wireless Ad Hoc Networks with Unidirectional Links”, *ACM/Kluwer Wireless Networks*, Vol.7, n. 6, pp. 617–625, 2001.
- [22] R. Ramanathan, R. Rosales-Hain, “Topology Control of Multihop Wireless Networks using Transmit Power Adjustment”, *Proc. IEEE Infocom 2000*, pp. 404–413, 2000.
- [23] V. Ramasubramanian, R. Chandra, D. Mosse, “Providing a Bidirectional Abstraction for Unidirectional Ad Hoc Networks”, *Proc. IEEE Infocom 02*, pp. 1258–1267, 2002.
- [24] V. Rodoplu, T.H. Meng, “Minimum Energy Mobile Wireless Networks”, *IEEE Journal Selected Areas in Comm.*, Vol. 17, n. 8, pp. 1333–1344, 1999.
- [25] P. Santi, D.M. Blough, “The Critical Transmitting Range for Connectivity in Sparse Wireless Ad Hoc Networks”, *IEEE Transactions on Mobile Computing*, Vol. 2, n. 1, pp. 1–15, January–March 2003.
- [26] W.Z. Song, Y. Wang, X.Y. Li, O. Frieder, “Localized Algorithms for Energy Efficient Topology in Wireless Ad Hoc Networks”, *Proc. ACM MobiHoc*, pp. 98–108, 2004.
- [27] P.J. Wan, C.W. Yi, “Asymptotical Critical Transmission Radius and Critical Neighbor Number for  $k$ -Connectivity in Wireless Ad Hoc Networks”, *Proc. ACM MobiHoc 04*, pp. 1–8, 2004.
- [28] R. Wattenhofer, L. Li, P. Bahl, Y. Wang, “Distributed Topology Control for Power Efficient Operation in Multihop Wireless Ad Hoc Networks”, *Proc. IEEE Infocom 2001*, pp. 1388–1397, 2001.
- [29] F. Xue, P.R. Kumar, “The Number of Neighbors Needed for Connectivity of Wireless Networks”, *Wireless Networks*, Vol 10, n. 2, pp. 169–181, 2004.

## Appendix – Proof of Th. 4

Let  $S_l$  denote the square of side  $l$  centered at the origin (i.e., the lower left corner of  $S_l$  is  $(-l/2, -l/2)$ ), where  $l$  is an arbitrary constant greater than 0. Consider the following scaling function  $f_l : S_l \mapsto S_1$ :

$$f_l(x, y) = \left(\frac{x}{l}, \frac{y}{l}\right)$$

The following facts are easy to prove:

**Fact 1.**  $f_l$  is a bijection.

**Fact 2.** If  $p_1, p_2 \in S_l$ , then  $d(p_1, p_2) = l \cdot d(f_l(p_1), f_l(p_2))$ .

**Fact 3.** Let  $S \subseteq S_l$  be a rectangular region with sides parallel to the axes. Then  $\text{area}(S) = l^2 \cdot \text{area}(f_l(S))$ .

**Fact 4.** If  $p_1, p_2, \dots, p_n$  are chosen independently and uniformly at random in  $S_l$ , then  $f_l(p_1), f_l(p_2), \dots, f_l(p_n)$  are independently and uniformly distributed in  $S_1$ . In particular, uniformity follows from Fact 3.

**Fact 5.** Let  $P = \{p_1, p_2, \dots, p_n\} \subseteq \mathbb{R}^2$  and let  $1 \leq k < n$ . For any  $p \in P$ , let  $N_{P,k}(p)$  denote the set of the  $k$  points in  $P$  closest to  $p$ . If  $P \subseteq S_l$ , then  $Q = f_l(P) \subseteq S_1$  and  $N_{Q,k}(f_l(p)) = f_l(N_{P,k}(p))$ . In words, the  $k$  closest neighbors of  $f_l(p)$  in  $Q$  are those in the image set  $f_l(N_{P,k}(p))$  of the  $k$  closest neighbors of  $p$  in  $P$ . Note that this follows immediately from Fact 2. The reverse implication holds as well.

For any  $P = \{p_1, p_2, \dots, p_n\} \subseteq \mathbb{R}^2$ , let  $G_k(P)$  denote the  $k$ -neighbors graph over  $P$ . Similarly, we can define the symmetric super and sub-graph of  $G_k(P)$ , denoted  $G_k^+(P)$  and  $G_k^-(P)$ , respectively.

**Lemma 3.** *Let  $P = \{p_1, \dots, p_n\}$  be an arbitrary set of points in  $S_l$ . Then  $G_k(P)$  is strongly connected if and only if  $G_k(Q)$  is strongly connected, where  $Q = f_l(P) \subseteq S_1$ .*

*Proof.* Suppose  $G_k(P)$  is strongly connected. Then, for any  $p, q \in P$ ,  $q$  is reachable from  $p$ , i.e., there is a sequence of vertices  $p = p_1, p_2, \dots, p_t = q$  such that  $(p_i, p_{i+1})$  is an edge in  $G_k(P)$ , for  $i = 1, \dots, t-1$ . By definition, this means that  $p_{i+1} \in N_{P,k}(p_i)$ . By Fact 5, it follows that  $f_l(p_{i+1}) \in N_{Q,k}(f_l(p_i))$  for any  $i$ , which means that  $(f_l(p_i), f_l(p_{i+1}))$  is an edge of  $G_k(Q)$  for  $i = 1, \dots, t-1$ . Thus,  $f_l(q)$  is reachable from  $f_l(p)$  in  $G_k(Q)$ . Since  $p$  and  $q$  are arbitrary nodes, it follows that  $G_k(Q)$  is strongly connected. The reverse implication easily follows by exchanging the roles of  $P$  and  $Q$ .  $\square$

Note that Lemma 3 can be easily extended to  $G_k^+(P)$  and  $G_k^-(P)$ .

We make use of Lemma 3 to prove that the distribution of connected  $k$ -neighbors graphs with vertex set in  $S_l$  does not depend on  $l$ .

**Lemma 4.** *Let  $1 \leq k < n$  and  $l > 0$ . The probability that the  $k$ -neighbors graph  $G_k(P)$  is connected, where  $P = \{p_1, \dots, p_n\} \subseteq S_l$ , is independent of  $l$ , provided that  $p_1, \dots, p_n$  are chosen independently and uniformly at random.*

*Proof.* We shall prove that the probability that  $G_k(P)$  is connected equals the probability that the  $k$ -neighbors graph is connected when the vertices are chosen independently and uniformly at random in  $S_1$ . For  $i = 1, \dots, n$ , let  $\partial P_i \subseteq S_l$  denote a sufficiently small region around  $p_i$ , such that  $\partial P_i \cap \partial P_j = \emptyset$  if  $i \neq j$ . Here, sufficiently small means that, for any  $P = \{p_1, \dots, p_n\}$ , the  $k$ -neighbors graph  $G_k(P)$  does not change if  $p_i$  moves in  $\partial P_i$ , for

$i = 1, \dots, n$ . Analogously, let  $Q = \{f_l(p_1) = q_1, \dots, f_l(p_n) = q_n\} \subseteq S_1$ , and  $\partial Q_i = f_l(\partial P_i)$ . By Fact 2 we can conclude that the  $k$ -neighbors graph  $G_k(Q)$  does not change as  $q_i$  moves in  $\partial Q_i$ . Furthermore, by Lemma 3 we have:

$$\begin{aligned} \Pr\{G_k(P) \text{ is connected} \mid p_i \in \partial P_i, i = 1, \dots, n\} &= \\ \Pr\{G_k(Q) \text{ is connected} \mid q_i \in \partial Q_i, i = 1, \dots, n\} & \end{aligned} \quad (2)$$

(It is immediate to see that the above conditional probability is 0 or 1.)

Now, suppose that  $p_1, \dots, p_n$  are chosen independently and uniformly at random in  $S_l$ . From the definition of conditional probability, and the independence assumption, we have

$$\begin{aligned} \Pr\{G_k(P) \text{ is connected} \wedge \text{all } p_i \in \partial P_i\} &= \chi \cdot \Pr\{\text{all } p_i \in \partial P_i\} \\ &= \chi \cdot \prod_{i=1}^n \Pr\{p_i \in \partial P_i\}, \end{aligned} \quad (3)$$

where  $\chi$  denotes the value of (2), while the shorthand “all  $p_i \in \partial P_i$ ” stands for “ $p_i \in \partial P_i, i = 1, \dots, n$ ”. Since by Fact 4  $q_1, \dots, q_n$  are also independently and uniformly distributed in  $S_1$ , we have:

$$\begin{aligned} \Pr\{G_k(Q) \text{ is connected} \wedge \text{all } q_i \in \partial Q_i\} &= \chi \cdot \Pr\{\text{all } q_i \in \partial Q_i\} \\ &= \chi \cdot \prod_{i=1}^n \Pr\{q_i \in \partial Q_i\}. \end{aligned} \quad (4)$$

The final step is to compute the probability of connectivity by integrating (3) and (4):

$$\begin{aligned} \Pr\{G_k(P) \text{ is connected}\} &= \int_{S_l} \dots \int_{S_l} \Pr\{G_k(P) \text{ is connected} \wedge \text{all } p_i \in \partial P_i\} \\ &= \int_{S_l} \dots \int_{S_l} \chi \cdot \Pr\{\text{all } p_i \in \partial P_i\} = \chi \cdot \prod_{i=1}^n \int_{S_l} \Pr\{p_i \in \partial P_i\} = \chi \cdot \prod_{i=1}^n \int_{S_l} \frac{\partial P_i}{l^2} \\ &= \chi \cdot \prod_{i=1}^n \int_{S_1} \partial Q_i = \Pr\{G_k(Q) \text{ is connected}\}, \end{aligned}$$

where the last equality follows clearly by applying the very same transformations in reverse order.  $\square$

As in the case of Lemma 3, the result of Lemma 4 can be easily extended to the symmetric super and sub-graph of the  $k$ -neighbors graph.

The proof of the theorem follows immediately from Theorems 2 and 3 and from Lemma 4.