

Probabilistic Algorithms for Cost-based Integrated MAC and Routing in Wireless Sensor Networks

Michele Rossi
Dept. of Engineering, University of Ferrara
via Saragat 1, 44100 Ferrara, Italy
mrossi@ing.unife.it

Michele Zorzi
Dept. of Inf. Engineering, University of Padova
via Gradenigo 6/B, 35131 Padova, Italy
zorzi@ing.unife.it

ABSTRACT

In this paper we propose a new integrated MAC/routing algorithm for CSMA channel access based wireless sensor networks (WSNs). In modern joint MAC/routing solutions for WSNs, forwarding decisions are often made based on metrics related to the advancements toward a specific destination (sink) and this is usually achieved through geographical routing and/or local indicators (virtual topologies) of the direction to be followed to get to the sink. In these settings, an important problem to be solved consists of providing extremely efficient next hop selection procedures, where relay nodes should be selected in a short time, in an energy efficient manner and promoting the nodes leading to positive geographical advancements toward the sink. In this work, we generalize the concept of node energy by referring to generic node costs, that we introduce here to represent the suitability of a given sensor node to be selected for data forwarding. Costs may therefore be associated with node energies, queue states (network congestion) as well as link qualities. Further, we propose to jointly exploit local routing rules (coordinates or virtual topologies) and node costs right into the channel access phase. In our framework, potential relay nodes contend for the channel based on a cost-based probabilistic approach where all these aspects are jointly considered and the final goal is to promote the selection of cost-efficient paths.

Categories and Subject Descriptors

C.2.1 [Network Architecture and Design]: Wireless Communication; C.2.2 [Computer-Communication Networks]: Network Protocols; C.4 [Performance of Systems]: Performance Attributes; I.1.6 [Computing Methodologies]: Simulation and Modeling

General Terms

Algorithms, Design, Modeling, Performance

Keywords

Wireless Sensor Networks, Routing, Media Access Protocols, Cost Efficient Communication, Cross-layer Design

1. INTRODUCTION

A wireless sensor network (WSN) consists of a number of sensors spread across a geographical area. Each sensor has wireless communication capability and some level of intelligence for signal processing and networking. The aim of the WSN is to gather, process and finally deliver environmental measurements to some central unit (*sink*). However, it shall be observed that the communication in WSNs is subject to stronger constraints than in wireless Ad Hoc networks as WSNs consist of low cost devices with very limited computation and storage capabilities. In addition, the requirements to be respected by WSN algorithms are to promote energy efficient solutions at every level of the protocol stack so as to prolong the network lifetime [1]. For these reasons, algorithms for WSNs must be *simple*, as sensor nodes can not deal with complex operations, and *energy efficient*, as the WSN should “survive” unattended as long as possible. A wide literature [2–12] was dedicated to the study of efficient solutions to these problem and, in particular, to propose efficient routing algorithms. One of the chief routing methodologies proposed so far [3, 4, 7, 9, 12] consists of the exploitation of geographical coordinates, where every node knows its position in the network as well as the position of the *sink*; geographical knowledge enables on-line routing schemes where the decision of the relay node can be accomplished based on geographical advancements toward the destination. In geographical routing, a sensible solution to perform data forwarding [13, 14] is to select those nodes leading to the maximum geographical advancements toward the sink; this technique is also referred to as Most Forward within Radius (MFR) [13]. Other alternatives are given by Greedy Routing schemes (GRS) [7, 15], where the packet is forwarded to the closest nodes to the destination. However, in addition to the minimization of the distance to the sink, other goals, such as the maximization of the node residual energies, should also be taken into account. This may be achieved e.g. by exploiting node power off states [5, 7]. The above schemes are examples of *localized algorithms*, where each node forwards packets based on the state of the nodes in its first order neighborhood. This is pivotal to avoid the resource wastage deriving from the dissemination of network-wide connectivity/cost messages. In [12] the authors introduce the concept of *partial topology knowledge forwarding*. In their framework, they determine the optimal *knowledge range* (radio range) in order to make efficient localized forwarding decisions. The aim is to improve the network view of a given node in order to approach globally optimal solutions by means of “controlled” local views of the network. The main focus in [12] is on the energy expenditure. Here, we stress that geographic routing is inherently affected by the following two problems: i) every node must be aware of its own position and of the position of the sink and ii) situations may occur where there are no nodes leading to a positive advancement toward the destination (*connectivity hole* problem [3, 4, 9]). Moreover, geographical coordinates are often obtained by costly operations or

by dedicated hardware (GPS) that, for certain applications, may be too expensive to have at every node. Further, the lack of connectivity calls for new algorithmic solutions [3, 4, 9] that likely lead to sub-optimal paths and only with probabilistic delivery guarantees. As a possible solution to i) and ii), in [10, 11] forwarding decisions are made based on virtual topologies. In addition, path optimality is achieved by propagating communication costs so as to build a *minimum cost field*. Once such a cost field has been set up, every node can communicate with the most cost efficient node within range that leads to a positive advancement toward the destination.

As observed above, in WSNs the maximization of system efficiency is a key goal. More than this, system wide efficiency shall be achieved at every layer of the protocol stack. In fact, at a given instant, the best node in terms of physical layer-related conditions (link quality) may be the worst in terms of network related metrics as its internal queue may be congested or its battery exhausted. Hence, we clearly have physical related metrics (e.g., link quality) and network-related metrics (e.g., residual energies) that coexist together and that *must be considered jointly*. These arguments lead us to the concept of cross-layer design [16, 17] as a tool for merging the requirements at every level of the protocol stack and seek for global efficiency. Driven by these reasonings, in this work we propose to integrate routing decision rules right into the channel access phase (MAC). As in geographic routing, we still need some topology related information to select the best relay node for the current transmission. In this respect, we focus on virtual topologies [10] [11], however, we stress that our channel access methodology can be adapted through minor modifications to geographical routing as well. Previous research considering a cross-layer design of routing and channel access can be found in [7], where the authors introduce the concept of receiver driven contention and present a practical and effective algorithm for the relay node selection, which exploits a heuristic partitioning of the forwarding area. The present paper consists of a generalization of previous approaches [7] [12] as we substitute geographical forwarding regions with probabilistic forwarding regions which are built considering node-specific/link-specific metrics as well as advancements toward the destination. These quantities are encoded here in a single node parameter which is referred to as “cost”. For the channel contention, we use analytically derived channel access probabilities that are designed to promote the lowest cost node within range. Hence, our approach deals with a probabilistic selection of the nodes within coverage that satisfy certain requirements.

The remainder of this paper is organized as follows. In Section 2 we introduce the network and cost models. In Section 3, we discuss routing rules for hop count (HC) virtual topologies, by formalizing routing over virtual topologies as a sequential decision problem and presenting a possible on-line localized HC strategy. Sections 4 and 5 present a probabilistic method to derive channel access probabilities and design a MAC scheme to be coupled with localized routing solutions. These sections report the most important contribution of the paper. Section 6 presents the performance evaluation and finally Section 7 concludes our work.

2. NETWORK MODEL

The network is modeled as a weighted graph $\mathcal{G} = (\mathcal{M}, \mathcal{A})$, where \mathcal{M} is the set of nodes and \mathcal{A} is the set of links between nodes. Among the $m = |\mathcal{M}|$ nodes in \mathcal{M} , we consider a processing unit referred to as *sink* and $m - 1$ nodes whose function is to generate traffic and forward packets towards the sink using a multi-hop routing technique. \mathcal{A} is a set of ordered pairs (i, j) , where $i, j \in \mathcal{M}$. (i, j) is referred to as the link connecting node i to node j . (i, j)

exists iff node j can successfully receive the messages transmitted by node i ($i \rightarrow j$). In our research we try to keep the communication/connectivity model as general as possible, without making any specific assumption on physical channel model, modulation and coding techniques. For this reason, any further specification on these aspects would not affect the validity of the results presented next. The analysis is based on the concept of *neighboring sets*, i.e., on sets of nodes within coverage of a given node and at a given time instant. We stress that neighboring sets may dynamically vary between subsequent forwarding actions, depending on the network configuration, channel behavior, mobility and, among other factors, node sleeping cycles [5, 7]. It would therefore be infeasible to derive these sets once for all, e.g., at the beginning of network operations, whereas it is reasonable to obtain neighboring sets *on-demand* and when the forwarding decision has to be actually taken. Each arc $(i, j) \in \mathcal{A}$ is characterized by a bounded normalized cost $c_{ij} \in [0, 1]$, which depends on the resources that are needed to transmit a message from node i to node j . Such a function could be related, for instance, to the energy required to transmit a single information bit, but many other factors can also be taken into consideration, e.g., the quality associated with link (i, j) , the congestion level at node j (e.g., the state of the queue at node j), the node failure probability or the residual energy of the nodes (if they have a limited energy reserve). Here, in order to keep our framework as general as possible, we do not propose a specific model to determine such costs. A *path* from node i to node d is defined as an ordered list of nodes, i.e., $\mathcal{P} = \{s, r_1, r_2, \dots, r_{n-1}, d\}$, where nodes s and d are referred to as the *source* and the *destination* node, respectively; nodes $r_j, j \in \{1, 2, \dots, n-1\}$ in \mathcal{P} are referred to as relay nodes. In our analysis, we only consider all loop-free oriented paths connecting node s to node d . The end-to-end cost $C(\mathcal{P})$ of path \mathcal{P} (from $s \rightsquigarrow d$) is obtained as the sum of the costs associated with every link in the path¹

$$C(\mathcal{P}) = c_{sr_1} + \sum_{i=1}^{n-2} c_{r_i r_{i+1}} + c_{r_{n-1} d} \quad (1)$$

The choice of an additive cost function as the path cost criterion is reasonable as additive metrics arise in many settings. For example, end-to-end delay, delay jitter, maximum total residual energy and reliability all correspond to the sum of link weights. In the present work, we assume that the cost c_{ij} of the link between i and j does not depend on node i , that is, $c_{ij} = c_j, \forall j \in \mathcal{M}$. This simplifying assumption is reasonable, for instance, in a scenario where all nodes transmit with the same constant power, but may of course be removed in future research. Moreover, even when nodes do transmit with different powers this assumption may be reasonable to minimize any *transmission power independent metric*. One example in this sense is represented by the need to avoid congestion, where queue states could be used at each node to calculate congestion costs.

3. ROUTING OVER VIRTUAL TOPOLOGIES

In the following subsections we present a possible framework to perform routing over multi-hop WSNs based on virtual topologies. This framework is reported here as an example to be coupled with the channel access solutions that will be presented in Section 4.

3.1 Hop Count Virtual Topologies

For the virtual topology we adopt the following gradient algorithm (similar to the one proposed in [11]). The procedure has to be re-executed only when network nodes move. In the case of sensor

¹Under the hypothesis of additive cost function, see [18].

networks characterized by fixed nodes, it has to be executed only once in the deployment start-up phase

- The sink node (SN) initially broadcasts a hop count packet (*hc_pkt*) with Hop Count (HC) value 1, all the sensors that receive this packet store this value.
- Each node that receives a *hc_pkt*, say with hop count i , broadcasts a new *hc_pkt* with hop count $i + 1$. The procedure is repeated recursively until all nodes in the network have received and forwarded a *hc_pkt*.
- If a node receives more than one *hc_pkt*, the one containing the *lowest hop count value* is considered to select the hop count value (HC) for the current node.

In the scenario we have in mind, nodes may be stationary, are densely deployed, and can periodically enter sleep mode thereby providing a random topology. We also consider that each sensor can only transmit using a fixed power, i.e., no power control is accounted for. Moreover, for each node $i \in \mathcal{M}$, we define the set \mathcal{N}_i as the set containing all the neighboring nodes of node i . We further define $\mathcal{N}_i(n)$, $\mathcal{N}_i(n-1)$ and $\mathcal{N}_i(n+1)$, $n \in \mathbb{N}^+$ as the sets of neighbors of node i with hop count (HC) equal to n , $n-1$ and $n+1$, respectively, where $\mathcal{N}_i = \mathcal{N}_i(n-1) \cup \mathcal{N}_i(n) \cup \mathcal{N}_i(n+1)$.

3.2 Routing as a Sequential Decision Problem

We formulate the routing process as a sequential decision problem, where at every stage a node has to select a specific action, i.e., the best node to act as relay for the current packet. Our focus is on *on-line* routing schemes, where forwarding decisions are made based on local network views and on some statistical information regarding the second-order (two-hops away) neighborhood of the current node. With the term *local views*, we mean the knowledge of the costs of the nodes within radio range. For now, we consider this information as available through an ideal MAC and at no cost. In Section 5, these routing rules will be combined with a novel probabilistic MAC thereby providing a practical integrated routing/MAC scheme. We assume that the node occupied in the current forwarding stage is node $i \in \mathcal{M}$, that its hop count is $\text{HC}(i) = n$ and that the forwarding process is at stage, $t \in \mathbb{N}$, where time evolves by one unit at every decision step. The problem to be solved by the decision maker is therefore to decide which is the best node to be selected to act as relay among the nodes in sets $\mathcal{N}_i(n)$ and $\mathcal{N}_i(n-1)$. Nodes in set $\mathcal{N}_i(n+1)$ are excluded *a priori* since, in normal conditions, they do not lead to satisfactory solutions.² We refer to $j_{n-1}^t \in \mathcal{N}_i(n-1)$, $j_n^t \in \mathcal{N}_i(n)$ and to c_{n-1}^t , c_n^t as the minimum cost nodes³ in sets $\mathcal{N}_i(n)$ and $\mathcal{N}_i(n-1)$ and their associated costs, respectively. We define *forwarding cycle* as the sequence of steps between the stage where a node with hop count n is reached for the first time and the instant in which a node with hop count $n-1$ is eventually selected as relay. In other words, a cycle is the number of steps it takes the decision maker (packet) to advance one hop toward the destination. As said above, the main objective of the routing scheme is to minimize the total cost of the final selected path, which is computed as in Eq. (1). In particular, the whole path can be decomposed as a disjoint sequence of n

²This has been verified by extensive simulation and is also supported by previous studies [7].

³In the case where there are multiple nodes with the same minimum cost in one of the two sets, we randomly pick any one of them as they are, by definition, equivalent.

```

d ← sink node
s ← source node
Ctotmin ← +∞
Cacc ← 0
T ← ∅
repeat
  i ← current node
  node jn-1 s.t. cn-1 = minz ∈ Ni(n-1) {cz}
  node jn s.t. cn = minz ∈ Ni(n) {cz}
  Ctotmin ← min{Ctotmin, Cacc + cn-1}

  if (Ctotmin - (Cacc + cn) ≤ ε) or (jn ∈ T) then
    - select next hop j ← jn-1
    - Cacc ← 0
    - Ctotmin ← +∞
  else
    - select next hop j ← jn
    - Cacc ← Cacc + cn
  end

  Update tabu list T:
  begin
    - T ← T ∪ j
    - Delete nodes with age ≥ tabulen from T
  end
until i = d;

```

Algorithm 1: Statistically-Assisted greedy Routing Algorithm (SARA). A tabu list T is used to prevent cycles or ping-ponging between nodes at the same hop count distance.

forwarding cycles⁴, i.e., its cost equals the summation of the costs incurred in each forwarding cycle. Further, if we assume that the forwarding cycles are statistically independent, the minimization of the expected cost associated with the full path is obtained by minimizing the expectation of each term taken separately. In this case, devising a scheme to minimize the path cost is equivalent to devising a procedure to minimize the expected cost of each cycle. The analytical framework we developed for this purpose is discussed in the following Section 3.3. As the major point of the paper consists of the proposal of a channel access methodology to be coupled with localized routing techniques, we keep the discussion in the next section at an intuitive level. A rigorous treatment of the matter can be found in [19].

3.3 Hop Count Routing Policies

Consider a generic forwarding step, $t \in \mathbb{N}$, and consider that the packet at time t is at node i with $\text{HC}(i) = n$ and that time 0 corresponds to the instant where the current *forwarding cycle* has started. At time t the *decision maker* (packet) has to choose a forwarding action, i.e., whether the packet is to be forwarded to node j_{n-1}^t or to j_n^t . We define the action set and the decision maker's current state as $\mathcal{A}_t = \{a_{n-1}^t = j_{n-1}^t, a_n^t = j_n^t\}$ and $X_t = (c_n^0, c_{n-1}^0, c_n^1, c_{n-1}^1, \dots, c_n^t, c_{n-1}^t)$, respectively. Moreover, we assume that if action $a(t) \in \mathcal{A}_t$ is chosen when in state X_t , $t \geq 0$, a cost $C(X_t, a(t)) \geq 0$ is incurred. Moreover, if at time $t \geq 0$ decision a_{n-1}^t is made, then node j_{n-1}^t is selected and the cycle ends with a total cost $C_{tot}(t)$

$$C_{tot}(t) = C_{par}(t) + c_{n-1}^t \quad (2)$$

⁴When the starting node $i \in \mathcal{M}$ has $\text{HC}(i) = n$.

where $C_{par}(t)$ is obtained as

$$C_{par}(t) = \begin{cases} 0 & t = 0 \\ \sum_{i=0}^{t-1} c_n^i & t \geq 1 \end{cases} \quad (3)$$

On the other hand, if decision a_n^t is taken, the cycle is continued toward node j_n^t with an accumulated partial cost $C_{par}(t+1)$. Observe that when $C_{par}(t+1) \geq C_{tot}(t)$ there is no point in further searching for a better path and the cycle should end. In this respect, note that if the current time is t , the two quantities c_{n-1}^t and c_n^t can be easily obtained as they are associated with two in-range nodes. In this sense they can be regarded as known quantities. Moreover, $C_{tot}(t)$ is also known as it is calculated as a function of the costs incurred from the beginning of the current forwarding cycle. For what concerns $C_{par}(t+1)$, we instead observe that its value can not be known at time t as it depends on c_{n-1}^{t+1} , which is associated with an out of range and not yet reached node. However, as it will be explained next, in order to implement a good routing rule, we may account for a statistical characterization of c_{n-1}^{t+1} . This will lead us to a decision (routing) criterion which is based on statistical expectations rather than on actual values. The minimum cost of the paths encountered by the decision maker up to and including time t is evaluated as

$$C_{tot}^{min}(t) = \min_{0 \leq k \leq t} \left\{ C_{tot}(k) \right\} \quad (4)$$

At every decision stage t , the decision maker can keep track of the previously encountered costs $\{c_{n-1}^0, c_n^0, c_{n-1}^1, c_n^1, \dots, c_{n-1}^t, c_n^t\}$ by therefore evaluating the minimum cost of all paths encountered so far $C_{tot}^{min}(t)$ (Eq. (4)). Now, if the optimality criterion is to *select the minimum cost path*, the corresponding one-stage policy obeys the following stopping set [19] [20]

$$\mathcal{B} = \left\{ X_t : C_{tot}^{min}(t) - C_{par}(t+1) \leq \mathcal{E} \right\} \quad (5)$$

where $C_{par}(t+1)$ and $C_{tot}^{min}(t)$ are defined in Eqs. (3) and (4), respectively. \mathcal{E} is defined as

$$\mathcal{E} = E[c_{n-1}^{t+1}] = \int_0^1 c_{n-1}^{t+1} dF_{min}(c_{n-1}^{t+1}) \quad (6)$$

and represents the expected minimum cost among nodes with hop count $n-1$ at stage $t+1$ and $F_{min}(\cdot)$ is the minimum cost cdf. The aim of the stopping rule dictated by Eq. (5) is to continue the cycle until the expected cost of the path at the next step is higher than the minimum path cost encountered so far. It can be proven [19] that the policy dictated by set \mathcal{B} gives the globally optimal behavior for the decision maker, i.e., among all feasible policies exploiting the available first- and second-order cost information⁵ it is the policy leading to the lowest long-term expected cost [20].

The routing scheme defined by this policy is named Statistically Assisted Routing Algorithm (**SARA**) as it exploits two-hop neighborhood statistical measures (\mathcal{E}) to carry out the relay selection. The full version of the SARA routing scheme, which is a straightforward implementation of Eq. (5), is reported in Alg. 1. Observe that we also use a tabu list in order to avoid ping-ponging between nodes placed at the same hop count distance, i.e., a node with hop count n which is selected to forward a given packet at time t can not be re-elected as the relay for the same packet for *tabulen* subsequent forwarding actions. In all our simulations, we found this

⁵First-order minimum costs (c_n^t, c_{n-1}^t) and second-order minimum cost estimate $\mathcal{E} = E[c_{n-1}^{t+1}]$.

simple strategy to be very effective in completely avoiding routing loops.

4. A NOVEL DISTRIBUTED CHANNEL ACCESS TECHNIQUE

The aim of the following analysis is to propose and validate an efficient channel access methodology to be coupled with the routing technique discussed above. The main goal is to realize a simple procedure to gather the information related to minimum cost nodes in sets $\mathcal{N}_i(n)$ and $\mathcal{N}_i(n-1)$. It is worth observing that such a procedure should be distributed, i.e., the decision on whether a node has to access the channel should be made independently at every node and exploiting a minimal amount of information regarding the state of neighboring devices. Further, the algorithm should be efficient in the sense that: i) a limited number of contention rounds should be required to eventually establish a winner and ii) the winner of the contention should satisfy certain desirable properties, i.e., its cost should be sufficiently close to the minimum in its set. In the following section we propose a suitable approach to compute the node channel access probabilities, by qualitatively and quantitatively showing the role of the cost correlation ρ among competing nodes.

4.1 Computation of Channel Access Probabilities

As explained above, in our routing framework we exploit the minimum cost information of nodes in sets $\mathcal{N}_i(n-1)$ and $\mathcal{N}_i(n)$, where i is the node which is currently occupied by the forwarding process. In order to implement our routing rules in a practical and efficient integrated MAC/routing scheme, we need a simple and effective procedure to find the minimum cost nodes in sets $\mathcal{N}_i(n-1)$ and $\mathcal{N}_i(n)$. Let us focus on the nodes in $\mathcal{N}_i(n-1)$ first. Node i first transmits a request message (REQ) packet containing the hop count number $n-1$. At this point, we are interested in devising a procedure to find the minimum cost node in set $\mathcal{N}_i(n-1)$: ideally, in fact, this should be the only node replying to the REQ. Moreover, we want to achieve this goal in a fully distributed manner, i.e., every node only knows its own cost and is not informed about the status of other nodes.

In order to illustrate the problems involved in solving this task, let us consider a generic set of N nodes \mathcal{S}_N and consider that all nodes in the set reply with the same channel access probability $P_a^j = P_a$, $\forall j \in \mathcal{S}_N$. Moreover, let us define the success probability P_{succ} as the probability that a single node replies to the REQ, whereas all the other nodes stay silent, $P_{succ} = NP_a(1-P_a)^{N-1}$. In such a case, it can be shown [21] that P_{succ} is maximized when $P_a = 1/N$ and its maximum value is given by $P_{succ}^{max} = (1-1/N)^{N-1}$. However, we shall observe that this simple access scheme does not provide a way of discriminating the minimum cost node in set \mathcal{S}_N . In fact, all nodes access the channel with the same probability, irrespective of their costs. Moreover, it is reasonable to think of a distributed scheme where such probabilities actually differ and are computed depending on node costs, i.e., $P_a^j = P_a^j(c_j)$, where $c_j \in [0, 1]$ is the cost of node $j \in \mathcal{S}_N$. In what follows, we devise appropriate functions and techniques to relate node costs to channel access probabilities. With the term *appropriate* we mean able to keep P_{succ} to a reasonably high value, while ensuring that the cost of the winner is sufficiently close to the minimum in set \mathcal{S}_N . In Section 4.2, we treat the case of independent costs. The more general correlated cost case is investigated in Section 4.3.

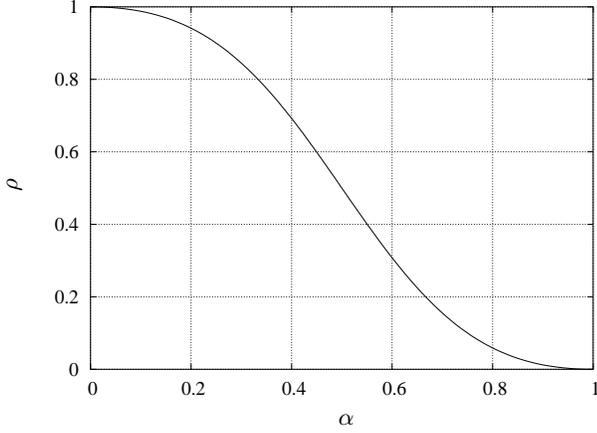


Figure 1: Cost correlation function ρ as a function of α .

4.2 Case I: Independent and Uniformly Distributed Costs

In order to devise a scheme for the minimum cost node discovery, we consider here the simplest but still important case where node costs are independent of each other. Consider the generic set \mathcal{S}_N of N nodes within coverage and consider that the cost associated with a given node in \mathcal{S}_N is independently drawn from a uniform distribution in $[0, 1]$. Moreover, consider a generic node $k \in \mathcal{S}_N$ whose cost is c_k . Owing to the independent cost assumption, we can evaluate the probability $P_{min}^k(c_k)$ of that node being the minimum cost node in set \mathcal{S}_N

$$P_{min}^k(c_k) = (1 - F_C(c_k))^{N-1} = (1 - c_k)^{N-1}, \quad c_k \in [0, 1] \quad (7)$$

where $F_C(c) = \text{Prob}\{C \leq c\}$ is the cumulative distribution function (cdf) of the cost. $P_{min}^k(c_k)$ is conditioned on the cost of node k (c_k) and corresponds to the probability that the remaining $N - 1$ nodes in \mathcal{S}_N have a cost higher than or equal to c_k . In fact, it could be reasonable to use $P_{min}^k(c_k)$ as the channel access probability P_a^k at node k . In order to clarify this assertion, let us write the MAC success probability P_{succ} as the probability of occurrence of the event where a single node out of the N in \mathcal{S}_N accesses the channel. If we choose $P_a^k = P_{min}^k(c_k)$, the expected value for this probability is given by

$$\bar{P}_{succ} = \underbrace{\int_0^1 \dots \int_0^1}_{N \text{ times}} \sum_{i=1}^N P_a^i \prod_{j \neq i} (1 - P_a^j) dc_1 \dots dc_N = \left(1 - \frac{1}{N}\right)^{N-1} \quad (8)$$

Hence, when costs are independent and uniformly distributed the choice $P_a^k = P_{min}^k(c_k)$ is optimal in the sense that, on average, P_{succ} equals the maximum success probability, that is obtained when all nodes access the channel with probability $P_a = 1/N$ (see previous Section 4.1). For non-uniform distributions, the above analysis can be easily extended and similar results are expected.

4.3 Case II: Correlated Costs

In what follows, we further complicate our cost model by taking into account the cost correlation between nodes in set \mathcal{S}_N and evaluating its impact on the MAC layer performance. As above, we consider a generic set \mathcal{S}_N of N nodes, where we refer to c_j , as the cost associated with node $j \in \mathcal{S}_N$. Moreover, in order to model the cost correlation among nodes, we assume that the r.v. C_j governing the cost of node j (c_j) is achieved by summing two

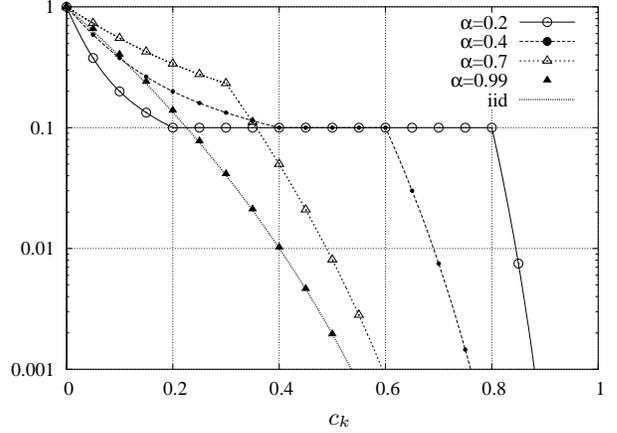


Figure 2: $P_{min}^k(c_k)$ as a function of c_k for $N = 10$ and $\alpha \in \{0.2, 0.4, 0.7, 0.99\}$. The independent cost case is also plotted for comparison.

r.v. \bar{C} and Γ_j as follows, $C_j = \bar{C} + \Gamma_j$, where $\bar{C} \in \mathcal{U}[0, 1]$ and $\Gamma_j \in \mathcal{U}[-\alpha\bar{c}, \alpha(1 - \bar{c})]$, $\mathcal{U}[a, b]$ is the uniform distribution in the interval $[a, b]$, $a, b \in \mathbb{R}$, $\alpha \in [0, 1]$ and \bar{c} is the actual value of the r.v. \bar{C} . Therefore, the cost of a generic node $j \in \mathcal{S}_N$ is given by a common part \bar{C} , which is equal for all nodes in \mathcal{S}_N , and an additive random displacement $\gamma_j \in [-\alpha\bar{c}, \alpha(1 - \bar{c})]$, which is independently picked for every node in the set but that depends on the actual value of the r.v. \bar{C} . \bar{c} in our model is used to represent the common cost component of nodes in \mathcal{S}_N . Clearly, the limiting cases $\alpha = 0$ and $\alpha = 1$ correspond to the fully correlated case, i.e., where all nodes in \mathcal{S}_N have the same cost \bar{c} , and to the independent case, i.e., where the costs of every pair of nodes in \mathcal{S}_N are uncorrelated, respectively. This is a simple model that we introduce to mathematically derive a precise relationship between the cost correlation ρ and $P_{min}^k(c_k)$. We observe that the model is in general not accurate for every network condition. However, it allows to find very useful functional forms for $P_{min}^k(c_k)$ that, as will be shown next, are also very effective under general cost models. We define the correlation coefficient between any two nodes $r, s \in \mathcal{S}_N$ as

$$\rho_{r,s} = \frac{E[C_r C_s] - E[C_r]E[C_s]}{\sigma_r \sigma_s} \quad (9)$$

where $\sigma_s^2 = E[(C_s - E[C_s])^2]$. By standard calculations $\rho_{r,s}$ can be derived as (Fig. 1)

$$\rho_{r,s} = \frac{(1 - \alpha)^2}{(1 - \alpha)^2 + \alpha^2} \quad (10)$$

Observe that the parameter α can be mapped as a function of $\rho = \rho_{r,s}$ as follows

$$\alpha = \begin{cases} \frac{\rho - 1 + \sqrt{\rho(1 - \rho)}}{2\rho - 1} & 0 \leq \rho < 1/2 \text{ and } 1/2 < \rho \leq 1 \\ 1/2 & \rho = 1/2 \end{cases} \quad (11)$$

Moreover, for a given node k with cost c_k , the probability that this node is the minimum cost node in set \mathcal{S}_N is given by the following Eqs. (12),(13). The full derivation for these probabilities is reported

in the Appendix.

$$P_{min}^k(c_k) \Big|_{0 < \alpha \leq 0.5} = \begin{cases} \frac{\alpha^N - (\alpha - c_k)^N}{N^{-1} c_k N \alpha^{N-1}} & 0 \leq c_k < \alpha \\ \frac{(1 - c_k)^{N-1}}{N \alpha^{N-1}} & \alpha \leq c_k \leq 1 - \alpha \\ \frac{(1 - c_k)^{N-1}}{N \alpha^{N-1}} & 1 - \alpha < c_k \leq 1 \end{cases} \quad (12)$$

$$P_{min}^k(c_k) \Big|_{0.5 < \alpha < 1} = \begin{cases} \frac{\alpha^N - (\alpha - c_k)^N}{c_k N \alpha^{N-1}} & 0 \leq c_k < 1 - \alpha \\ \frac{(1 - c_k)^N - (\alpha - c_k)^N}{(1 - \alpha) N \alpha^{N-1}} & 1 - \alpha \leq c_k \leq \alpha \\ \frac{(1 - c_k)^{N-1}}{N \alpha^{N-1}} & \alpha < c_k \leq 1 \end{cases} \quad (13)$$

For the limiting cases $\alpha = 0$ and $\alpha = 1$, $P_{min}^k(c_k)$ is defined as

$$P_{min}^k(c_k) \Big|_{\alpha=0} = \frac{1}{N}, \quad P_{min}^k(c_k) \Big|_{\alpha=1} = (1 - c_k)^{N-1}. \quad (14)$$

Observe that $P_{min}^k(c_k)$ is continuous in c_k and that in Eqs.(12),(13), $\lim_{c_k \rightarrow 0} P_{min}^k(c_k) = 1$ and $P_{min}^k(1) = 0$, i.e., a node with $c_k = 0$ is with probability one among the minimum cost nodes in \mathcal{S}_N , whereas as $c_k = 1$, the probability of being the minimum cost node is zero. Eqs.(12),(13) are therefore consistent with the notion of minimum cost node. Fig. 2 reports the above probability $P_{min}^k(c_k)$ as a function of the node cost c_k for some significant values of α . For the sake of completeness, the curve representing the independent case (iid, i.e., $\rho = 0$ or $\alpha = 1$) is also depicted. As a first observation, note that as the cost correlation (ρ) tends to 1 ($\alpha \rightarrow 0$), $P_{min}^k(c_k)$ converges to $1/N$, i.e., the optimal access probability in the fully correlated case. With a decreasing ρ , instead, the curve still assumes the value $1/N$ within the interval $c_k \in [\alpha, 1 - \alpha]$ (see Eq. (12)), whereas the probability of being the minimum cost node quickly goes to zero as the cost c_k increases beyond $(1 - \alpha)$ and tends to 1 as $c_k \rightarrow 0$. This behavior continues for a decreasing ρ (increasing α), up to the breaking point $\rho = \alpha = 0.5$, where the portion of the curve with value $1/N$ collapses to a single point. Further, as ρ continues to decrease, the behavior of the curve changes according to Eq. (13), by getting closer to the fully independent ($\rho = 0$) situation. Fig. 2 highlights the substantial differences in $P_{min}^k(c_k)$ as a function of ρ .

5. INTEGRATED MAC/ROUTING SCHEME

In the following we propose an integrated MAC/routing schemes that we call **SARA-M**, where the selection rule in Eq. (5) is coupled with Eqs. (12),(13),(14)

1. Assume that the forwarding process is currently at node i with $\text{HC}(i) = n$ and that the quantities $C_{min}^{tot}(t)$ (Eq. (4)) and $C_{par}(t)$ (Eq. (3)) have been computed by node i , where t is the number of forwarding steps elapsed from the beginning of the current forwarding cycle.
2. Node i initiates a first contention phase by sending a $\text{REQ}(n - 1, \rho = 0, N, T)$ message to trigger a reply from all active nodes in set $\mathcal{N}_i(n - 1)$. N is an estimate for the number of nodes in $\mathcal{N}_i(n - 1)$, and ρ is the estimated cost correlation. The initial value for ρ is set to zero. The REQ also includes a tabu list \mathcal{T} containing the identifiers of the last *tabulen* visited nodes (see Alg. 1).
3. Every active node $j \in \mathcal{N}_i(n - 1)$ computes $P_a^j = P_a^j(c_j, \rho, N)$ as explained in section 4.1, where c_j is the cost of node j , ρ

and N are the estimates contained into the REQ. All nodes $j \in \mathcal{N}_i(n - 1)$ reply to the REQ with probability P_a^j . If $j \in \mathcal{T}$, P_a^j is forced to be zero. In its *reply* (REP) message every node includes its own identifier and its cost.

4. The following three events can occur: (1) *collision*: more than one node in $\mathcal{N}_i(n - 1)$ reply to the REQ and no reply can be received (2) *silence*: no nodes reply (3) *success*: only one node, say node j_{n-1}^* , replies to the REQ or, multiple nodes reply but the message from node j_{n-1}^* is still decodable due to the capture effect. If either (1) or (2) occurs, the sender updates $\rho \leftarrow \min(1, \rho + \Delta\rho)$ and sends a REQ including the new ρ , and the contention is continued from step 2. If case (3) occurs, node j_{n-1}^* is the winner of the contention and, at node i , $c_{n-1}^t \leftarrow c_{j_{n-1}^*}^t$. This last case ends the contention.
5. Node i initiates a second contention phase by sending a $\text{REQ}(n, \rho = 0, N, T)$ to trigger a reply from all active nodes in set $\mathcal{N}_i(n)$. This contention phase proceeds as the previous one (steps 2-4) with the only difference that in this case every node includes the quantity \mathcal{E} (Eq. (6)) in the REP. If the winner of this contention is node j_n^* , then $c_n^t \leftarrow c_{j_n^*}^t$.
6. The relay node is chosen according to Eq. (5), i.e., node j_{n-1}^* is selected if $(C_{tot}^{min}(t) - (C_{par}(t) + c_n^t) \leq \mathcal{E})$, otherwise j_n^* is selected.

Observe that starting the contention phase with $\rho = 0$ is equivalent to considering independently distributed costs. This makes sense as $\rho = 0$ (Eq. (7)) corresponds to the case maximizing the probability of electing the minimum cost node. Therefore, we start the contention phase by privileging the goodness of the solution found and subsequently, and in case of *collision* or *silence*⁶, we modify ρ to increase the *success* probability. As a further observation, we note that in the above scheme ρ is the only parameter that is adaptively modified, whereas N is kept constant. The main reason for that is due to the sensitivity of $P_a(\cdot)$ against N , which is considerably lower than its sensitivity against ρ .⁷ Moreover, we observe that at least an average for N can be reasonably measured or estimated through, e.g., statistical characterization of the underlying node topology. Further results in this sense are left for future research.

6. RESULTS

As a reference model for our performance evaluation, we consider a random topology network, where nodes are placed according to a planar Poisson process with normalized node density $\lambda_n = \lambda \pi R^2$, where λ is the average number of nodes per unit area, whereas R is the constant transmission range. We consider a unit disk connectivity model [22], i.e., two nodes can communicate iff their distance is lower than or equal to R . Observe that the schemes presented in this paper can work for any topology setting as node density λ and connectivity model just translate into different neighboring sets, which are obtained on the fly each time a packet is to be transmitted. In such settings, λ_n can be seen as the average number of nodes actually awake (or active) within coverage. The simple connectivity

⁶The silence probability is very high when costs are correlated and we assume them independent and this, if no countermeasure is taken, may lead to extremely long channel contention phases.

⁷We have verified this through simulation. The detailed results are not reported here due to space constraints. However, an idea of the sensitivity against N and ρ can be gained from Fig. 2, where the dependence on N is mainly given by the horizontal line $1/N$.

model in [22] is considered here as a benchmark for the performance evaluation and will be improved, e.g., accounting for channel fading phenomena, in future extensions of the present work. In the results presented in the following, we consider $\lambda_n = 15$, $R = 1$, all nodes are assumed to be randomly placed on a square area of $18R \times 18R$ and to have a perfect estimate of \mathcal{E} . A study of the effects of estimation errors on \mathcal{E} is left for future research. In subsection 6.1 we test our solution considering a single contention round. In subsection 6.2 we report the multi-hop performance of the integrated MAC/routing schemes presented in Section 5.

6.1 Performance of the Single Hop Node Selection Algorithm

In this section, we first focus on the performance of the proposed probabilistic MAC within a single contention round. In this case, we consider a set \mathcal{S}_N of N nodes. In the following, we assume to have a perfect knowledge of N , whereas the correlation estimate sent along with REQ messages is computed as in Section 5. For the node costs we consider the analytical model (AN) presented in Section 4.3. In addition, in order to test the goodness of our approach in a more general setting and to give evidence that it is not exclusively tailored to the simple model of Section 4.3, we also adopt the NORTA [23] [24] framework which is a widely used method for representing random vectors whose component random variables have arbitrary marginal distributions and correlation matrix. In particular, we consider the case of uniformly distributed costs but with the more general correlation structure as given in [23, 24]. In Figs. 3 and 4, we plot $\overline{\Delta c}$ and \overline{n} considering both the AN and the NORTA cost models. We focus on a single MAC contention phase, according to steps 2 through 4 of the algorithm in Section 5, where the objective of the sender is to find the minimum cost node in \mathcal{S}_N . We define c_{min} as $c_{min} = \min_{j \in \mathcal{S}_N} (c_j)$ and c_{winner} as the cost of the winner of the contention phase. Moreover, we define $\overline{\Delta c}$ and \overline{n} as the expected value of the difference $\Delta c = c_{winner} - c_{min}$ and of the number of rounds n to complete the contention, i.e., to eventually elect a winner, respectively. In Fig. 3, we report $\overline{\Delta c}$ as a function of the actual cost correlation ρ , considering $N = 10$ and $\Delta\rho \in \{0.05, 0.1, 0.2\}$. In this figure, we show the MAC performance accounting for the channel access functions in the perfectly correlated (1/N, Section 4.1), independent (IID, Section 4.2), and correlated (COR, Section 4.3) cost cases. As expected, IID is the best scheme in terms of goodness of the solution found ($\overline{\Delta c}$), whereas 1/N is the worst. It is worth noting that COR stays very close to IID by leading to very good performance. Furthermore, $\Delta\rho$ can be effectively exploited to tune the behavior of the COR scheme, i.e., to decrease $\overline{\Delta c}$. Under the same settings and assumptions, in Fig. 4 we report \overline{n} as a function of the actual cost correlation ρ . In this case, the ranking between IID and 1/N is reversed. In fact, the channel access probabilities in 1/N (1/N) are chosen to maximize the success probability P_{succ} , but ignoring node costs. As a consequence, 1/N performance is the best in terms of delay (\overline{n}) and the worst in terms of cost ($\overline{\Delta c}$). On the other hand, IID accounts for node costs, but regardless of P_{succ} . This explains why, within our framework, IID and 1/N may be seen as lower bounds for $\overline{\Delta c}$ and \overline{n} , respectively⁸. From Fig. 4, we also stress that COR stays sufficiently close to the lower bound (1/N) and that, once again, the $\Delta\rho$ parameter can be varied to decrease \overline{n} . As a last observation, we point out that the role of $\Delta\rho$ depends on the considered performance metric. That is, an increasing $\Delta\rho$ is beneficial for the delay ($\overline{n} \downarrow$) but has a negative impact on the goodness

⁸Observe that the exact lower bound for the delay metric in the multiple slot case is given by the access probability distribution in [21].

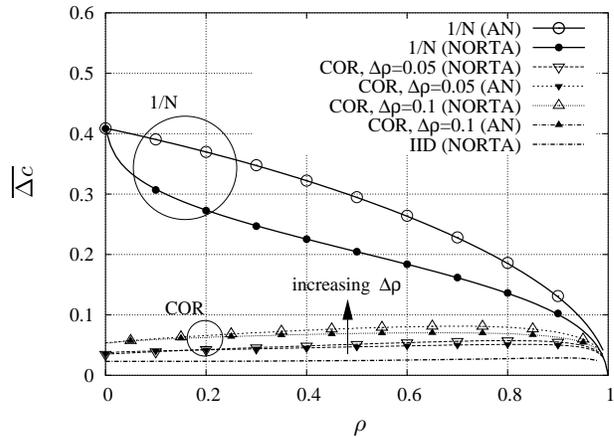


Figure 3: Quality of the solution found within a single contention phase. Number of nodes $N = 10$, $\Delta\rho \in \{0.05, 0.1, 0.2\}$.

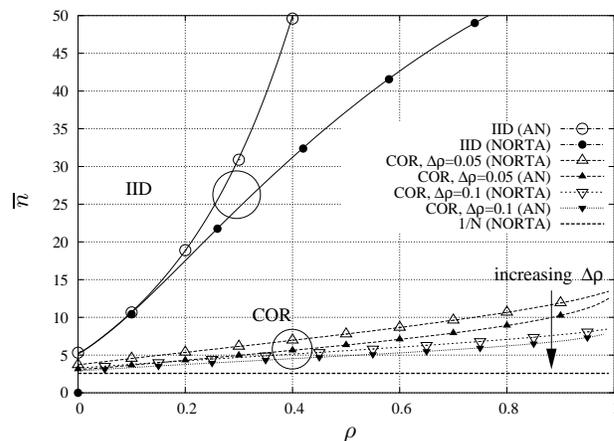


Figure 4: Average number of iterations to complete a single contention phase. Number of nodes $N = 10$, $\Delta\rho \in \{0.05, 0.1, 0.2\}$.

of the solution ($\overline{\Delta c} \uparrow$). A trade-off $\Delta\rho$ value is therefore to be found according to system/node requirements. From these graphs, it shall be observed that, even with a more general cost correlation model (NORTA), COR still performs very close to the optimal performance. While of course not providing any proof, these results provide strong evidence that our strategy is able to give very good performance without critical dependence on the cost correlation model, and therefore has wide applicability.

6.2 Multi Hop Performance of the Integrated MAC/Routing Schemes

In Fig. 5, we report the complementary cumulative distribution function (ccdf) of the difference between the cost (C , see Eq. (1)) of the path selected by the SARA on-line routing algorithm and the optimal cost solution (C_{opt}), which is found by solving an off-line optimization problem [25]. This cost distribution is computed by Monte Carlo simulation and is representative of the case where the hop count of the starting node is 8. In Fig. 5 we report the cost ccdf for the integrated MAC/routing SARA scheme, by considering the COR, IID, and 1/N MAC schemes. A “IDEAL” curve, where minimum cost nodes in the forwarding sets are obtained with probability

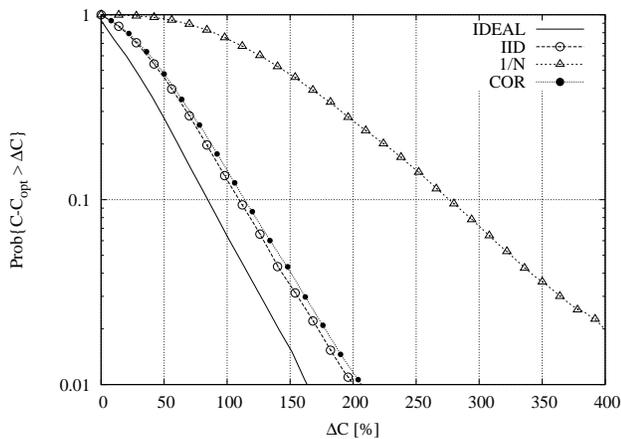


Figure 5: Cumulative distribution function of the path cost difference between the SARA-M integrated MAC/routing schemes and the optimal cost solution (C_{opt}). $\lambda_n = 15$, starting node hop count equal to 8, $\Delta\rho = 0.1$.

one and at no communication cost, is also plotted for comparison. Node costs are independently according to the $\mathcal{U}[0, 1]$ distribution. Observe that, even if costs in the network are picked independently, rather than the expected cost correlation, that in this case is zero, the key factor into the channel contention phase is the “instantaneous cost correlation” between the nodes in the forwarding sets. That is, the factor that affects the contention is the degree of “similarity” of the costs of the nodes at the beginning of a specific channel contention phase. Therefore, the above cost assignment model is sensible as the “instantaneous correlation” can assume any value and this means that we are testing our MAC under any possible situation. As highlighted in the figure, 1/N is largely suboptimal and should not be considered. On the other hand, COR when considered with the SARA routing algorithm gives satisfactory results as the degradation with respect to both IDEAL and IID is very small. It must be observed that, even if IID gives the best results in terms of cost, it is not a good candidate solution as its good cost performance is achieved at the expense of an extremely large number of contention periods (Fig. 4).

7. CONCLUSIONS

In this work we addressed routing and MAC algorithms for the efficient delivery of packets in wireless sensor and Ad Hoc networks. In the first part of the paper we briefly discussed novel *on-line* algorithms to efficiently route packets over virtual topologies, obtained by tracking hop count information at every node. The goal of these routing schemes is to minimize a generic cost function, that depends on the node costs encountered along the path and that can be related to, e.g., node residual energies or queue lengths. Subsequently, we proposed a new channel access methodology that takes into account the cost correlation between nodes. This strategy is specifically designed to efficiently identify, at a low communication cost, the minimum cost nodes within coverage and can be therefore efficiently coupled with any localized routing algorithm. As an example, we coupled the proposed probabilistic MAC with hop count routing schemes by obtaining very good results as the final integrated scheme performs satisfactorily with respect to both optimal solution and ideal scheme (collision free with complete local information). We observe that there are many issues that can be improved. For example, how the proposed scheme performs in the presence of malfunctioning nodes is still an open issue. Also, our

solution can be considerably improved by refining the contention procedure, and in particular its adaptation between subsequent contention rounds, by therefore further reducing the number of iterations needed to elect a relay node. All these issues are left for future research.

Acknowledgment

Part of this work has been supported by the Italian Ministry of Education, University and Research (MIUR) under the PATTERN project within the COFIN program.

8. REFERENCES

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “A Survey on Sensor Networks,” *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–116, Aug. 2002.
- [2] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva, “Directed Diffusion for Wireless Sensor Networking,” *IEEE/ACM Trans. on Networking*, vol. 1, pp. 2–16, Feb 2003.
- [3] I. Stojmenovic and X. Lin, “Loop-Free Hybrid Single-Path/Flooding Routing Algorithms with Guaranteed Delivery for Wireless Networks,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 12, no. 10, 2001.
- [4] B. Karp and H. T. Kung, “GPSR: Greedy Perimeter Stateless Routing for Wireless Networks,” in *Proceedings of IEEE/ACM Mobicom*, Boston, MA, Aug. 2000, pp. 243–254.
- [5] Y. Xu and J. Heidemann, “Geographic-Informed Energy Conservation for Ad Hoc Routing,” in *Proceedings of IEEE/ACM Mobicom*, Rome, Italy, July 2001, pp. 70–77.
- [6] P. Bose, P. Morin, I. Stojmenović, and J. Urrutia, “Routing with Guaranteed Delivery in Ad Hoc Wireless Networks,” *ACM Wireless Networks*, vol. 7, pp. 609–616, 2001.
- [7] M. Zorzi and R. R. Rao, “Geographic Random Forwarding (GeRaF) for Ad Hoc and Sensor Networks: Multihop Performance,” *IEEE Trans. on Mobile Computing*, vol. 2, pp. 337–348, Oct-Dec 2003.
- [8] G. Xing, C. Lu, R. Pless, and Q. Huang, “On Greedy Geographic Routing Algorithms in Sensing-Covered Networks,” in *Proceedings of the 5th ACM international symposium on Mobile ad hoc networking and computing*, Roppongi Hills, Tokyo, Japan, 2004, pp. 31–42.
- [9] F. Kuhn, R. Wattenhofer, Y. Zhang, and A. Zollinger, “Geometric Ad-Hoc Routing: of Theory and Practice,” in *Annual ACM Symposium on Principles of Distributed Computing*, Zurich, Switzerland, 2003, pp. 63–72.
- [10] F. Ye, A. Chen, S. Lu, and L. Zhang, “A Scalable Solution to Minimum Cost Forwarding in Large Sensor Networks,” in *IEEE 10th Int’l Conf. Computer Comm. and Networks*, 2001, pp. 304–309.
- [11] F. Ye, G. Zhong, S. Lu, and L. Zhang, “GRAdient Broadcast: A Robust Data Delivery Protocol for Large Scale Sensor Networks,” to *Appear in ACM Wireless Networks*, vol. 11, no. 2, Mar. 2005.
- [12] T. Melodia, D. Pompili, and F. I. Akyildiz, “Optimal Location Topology Knowledge for Energy Efficient Geographical Routing in Sensor Networks,” in *Proceedings of IEEE Infocom*, Hong Kong, P.R., China, Mar. 2004.

- [13] L. Kleinrock and J. Silvester, "Optimum Transmission Radii for Packet Radio Networks or why Six is a Magic Number," in *Proceedings of IEEE National Telecommunications Conference*, Birmingham, Alabama.
- [14] H. Takagi and L. Kleinrock, "Optimal Transmission Ranges for Randomly Distributed Radio Packet Terminals," *IEEE Trans. Commun.*, vol. 32, no. 3, pp. 246–257, 1984.
- [15] G. Finn, "Routing and Addressing Problems in Large Metropolitan Scale Internetworks," *ISI research report ISU/RR-87-180*, Mar. 1987.
- [16] Lang Tong and Vidyut Naware, "Signal Processing in Random Access," *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 29–39, Sept. 2004.
- [17] G. Dimic and N.D. Sidiropoulos and Ruifeng Zhang, "Medium access control - physical cross-layer design," *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 40–50, Sept. 2004.
- [18] Roch Guérin and Ariel Orda, "Computing Shortest Paths for Any Number of Hops," *IEEE/ACM Trans. Networking*, vol. 10, no. 5, pp. 613–620, Oct. 2002.
- [19] M. Rossi, M. Zorzi, and R. R. Rao, "Statistically Assisted Routing Algorithms (SARA) for Hop Count Based Forwarding in Ad Hoc and Wireless Sensor Networks," *Submitted to ACM Wireless Networks*.
- [20] S. M. Ross, *Introduction to Stochastic Dynamic Programming*. Academic Press, 1983.
- [21] Y. C. Tay, K. Jamieson, and H. Balakrishnan, "Collision-Minimizing CSMA and its Applications to Wireless Sensor Networks," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 1048–1057, Aug. 2004.
- [22] B. N. Clark, C. J. Colbourn, and D. S. Johnson, "Unit Disk Graphs," *Discrete Mathematics*, vol. 86, no. 1-3, pp. 165–177, Aug. 1991.
- [23] M. Cario and B. Nelson, "Modeling and Generating Random Vectors with Arbitrary Marginal Distribution and Correlation Matrix," in *Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois*, 1997.
- [24] S. Ghosh and S. G. Henderson, "Properties of the NORTA Method in Higher Dimensions," in *Proceedings of the 2002 Winter Simulation Conference*, San Diego, CA, US, Dec. 2002.
- [25] W.-K. Chen, *Theory of Nets: Flow in Networks*. John Wiley & Sons, 1990.

APPENDIX

A. DERIVATION OF COST-DEPENDENT ACCESS PROBABILITY FUNCTIONS

Consider the cost model introduced in Section 4.3 and consider a generic set of N nodes \mathcal{S}_N . In this Appendix we will derive the probability for the generic node $k \in \mathcal{S}_N$ to be the minimum cost node in the set (Eqs. (12)(13) and (14)). The actual cost at node k is $c_k = \bar{c} + \gamma_k$, where the common cost component (\bar{c}) is uniformly distributed in $[0, 1]$ and γ_k is uniformly distributed in

$[-\alpha\bar{c}, \alpha(1 - \bar{c})]$. If Γ_k is the r.v. associated with γ_k we have that $f_\Gamma(\gamma)$, the probability density function (pdf) of Γ_k is⁹

$$f_\Gamma(\gamma) = \begin{cases} \alpha^{-1} & \gamma \in [-\alpha\bar{c}, \alpha(1 - \bar{c})] \\ 0 & \text{elsewhere} \end{cases} \quad (15)$$

Moreover, considering the cost c_k as given, the pdf of \bar{c} given c_k , $f(\bar{c}|c_k)$ can be obtained via the Bayes rule as follows $f(\bar{c}|c_k) = f(c_k|\bar{c})f_{\bar{c}}(\bar{c})/f_C(c_k)$, where $f(c_k|\bar{c})$ is derived from Eq. (15) through a domain shift

$$f(c_k|\bar{c}) = \begin{cases} \alpha^{-1} & c_k \in [\bar{c} - \alpha\bar{c}, \bar{c} + \alpha(1 - \bar{c})] \\ 0 & \text{elsewhere} \end{cases} \quad (16)$$

$f_C(c_k)$, the pdf of the cost at node c_k is found via the following convolution integral

$$f_C(c_k) = \int_{-\infty}^{+\infty} f_{\bar{c}}(\bar{c}) f_\Gamma(c_k - \bar{c}) d\bar{c} \quad (17)$$

where the pdf associated with the common cost component $f_{\bar{c}}(\bar{c})$ is uniform in $[0, 1]$ and $f_\Gamma(\gamma_k)$ is defined as in Eq. (15). From the definition intervals of $f_{\bar{c}}(\cdot)$ and $f_\Gamma(\cdot)$, expressed as a function of \bar{c} (the common cost part), one can calculate the integral in Eq. (17) for 6 disjoint cases. These cases are reported in Fig. 6. In fact, the length of the definition interval of $f_\Gamma(\gamma_k)$ as a function of \bar{c} (the interval is $[(c_k - \alpha)/(1 - \alpha), c_k/(1 - \alpha)]$) is given by $\alpha/(1 - \alpha)$ and is smaller than or equal to 1 for $0 < \alpha \leq 0.5$ and larger than 1 when $0.5 < \alpha < 1$: this gives the two cases *a* and *b* in Fig. 6. Moreover, for each of these cases we have three possibilities, depending on the value of c_k . Hence, by solving the integral above, one achieves two different expressions for $f_C(c_k)$, according to the value of α . In particular, $f_C(c_k)$ is found for the two cases where $0 < \alpha \leq 0.5$ (case *a*) and $0.5 < \alpha < 1$ (case *b*) as follows

$$f_C(c_k) \Big|_{0 < \alpha \leq 0.5} = \begin{cases} \frac{c_k}{\alpha(1 - \alpha)} & 0 \leq c_k < \alpha \\ \frac{1}{1 - \alpha} & \alpha \leq c_k \leq 1 - \alpha \\ \frac{1 - c_k}{\alpha(1 - \alpha)} & 1 - \alpha < c_k \leq 1 \\ 0 & \text{elsewhere} \end{cases} \quad (18)$$

$$f_C(c_k) \Big|_{0.5 < \alpha < 1} = \begin{cases} \frac{c_k}{\alpha(1 - \alpha)} & 0 \leq c_k < 1 - \alpha \\ \alpha^{-1} & 1 - \alpha \leq c_k \leq \alpha \\ \frac{1 - c_k}{\alpha(1 - \alpha)} & \alpha < c_k \leq 1 \\ 0 & \text{elsewhere} \end{cases} \quad (19)$$

The conditional pdf $f(\bar{c}|c_k)$ can therefore be obtained as $f(\bar{c}|c_k) = f(c_k|\bar{c})f_{\bar{c}}(\bar{c})/f_C(c_k)$ that, can be written in a compact form as

$$f(\bar{c}|c_k) = \begin{cases} (\alpha f_C(c_k))^{-1} & \bar{c} \in \mathcal{I}_{\bar{c}} \\ 0 & \text{elsewhere} \end{cases} \quad (20)$$

where $f_C(c_k)$ is one of the two expressions above, according to the value of α . The interval $\mathcal{I}_{\bar{c}} = [\bar{c}_{min}, \bar{c}_{max}]$ is specified by

$$\bar{c}_{min} = \max\left(0, \frac{c_k - \alpha}{1 - \alpha}\right), \quad \bar{c}_{max} = \min\left(1, \frac{c_k}{1 - \alpha}\right). \quad (21)$$

⁹This pdf is conditioned on \bar{c} .

Case a : $0 < \alpha \leq 0.5$

Case b : $0.5 < \alpha < 1$

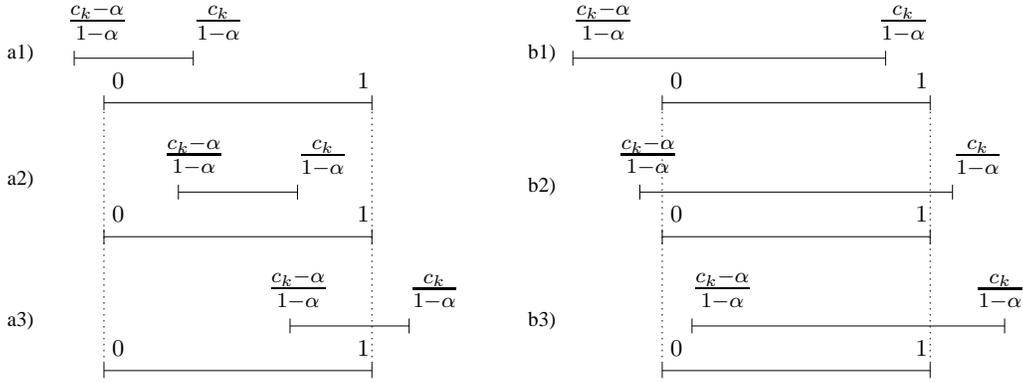


Figure 6: Diagram for the calculation of the convolution integral in Eq. (17). The intervals represent the values of \bar{c} for which $f(c_k|\bar{c})$ ($\bar{c} \in [(c_k - \alpha)/(1 - \alpha), c_k/(1 - \alpha)]$) and $f_{\bar{C}}(\bar{c})$ ($\bar{c} \in [0, 1]$) are different from zero.

Now, we refer to $P_{min}^k(c_k)$ as the probability that node k with cost c_k is the minimum cost node in \mathcal{S}_N . $P_{min}^k(c_k)$ is given by

$$P_{min}^k(c_k) = \int_{\bar{c} \in \mathcal{I}_{\bar{c}}} f(\bar{c}|c_k) (1 - F_{\Gamma}(c_k - \bar{c}))^{N-1} d\bar{c} \quad (22)$$

where $F_{\Gamma}(\cdot)$ is the complementary distribution function (cdf) of the r.v. Γ_k for the cost displacement at node k . The equation above returns the probability that all the remaining $N - 1$ nodes in \mathcal{S}_N have a cost which is higher than or equal to c_k (term $(1 - F_{\Gamma}(c_k - x))^{N-1}$), given that the cost of node k is c_k . The above integral must be evaluated considering the six cases above. For the sake of illustration, in the following we report the calculation for case a

a1) For a given value of $\alpha \in (0, 0.5]$, the integral in Eq. (22) is significant when $f(\bar{c}|c_k) > 0$. Moreover, $f(\bar{c}|c_k) > 0$ when both $f(c_k|\bar{c})$ and $f_{\bar{C}}(\bar{c})$ are larger than zero and this happens when $0 \leq \bar{c} < c_k/(1 - \alpha)$ (see case a1 in Fig.6). This gives the integration limits for Eq. (22). The pdf $f_C(c_k)$ that has to be considered to derive $f(\bar{c}|c_k)$ is given by the first line in Eq. (18). In fact, case a1 is equivalent to having $(c_k - \alpha)/(1 - \alpha)$, the inferior limit of the interval on the left-top side of Fig. 6, smaller than zero which translates to $0 \leq c_k < \alpha$, by therefore leading to the expression for $f_C(c_k) = c_k/(\alpha(1 - \alpha))$. Hence, $f(\bar{c}|c_k) = (1 - \alpha)/c_k$ (Eq. (20)), the cdf $F_{\Gamma}(\gamma) = (\gamma + \alpha\bar{c})/\alpha$ and Eq. (22) results as

$$\begin{aligned} & \int_0^{c_k/(1-\alpha)} \frac{(\alpha - c_k + \bar{c} - \alpha\bar{c})^{N-1} (1 - \alpha)}{c_k \alpha^{N-1}} d\bar{c} \\ &= \frac{1}{c_k \alpha^{N-1}} \int_{\alpha - c_k}^{\alpha} \xi^{N-1} d\xi = \\ &= \frac{\alpha^N - (\alpha - c_k)^N}{c_k N \alpha^{N-1}} \end{aligned} \quad (23)$$

and this gives the first line in Eq. (12) which holds for $0 \leq c_k < \alpha$, due to the condition on $f_C(c_k)$.

a2) For a given value of $\alpha \in (0, 0.5]$ and reasoning as above the interval to be considered for \bar{c} is $(c_k - \alpha)/(1 - \alpha) \leq \bar{c} \leq c_k/(1 - \alpha)$, the region for c_k ($f_C(c_k)$) is $\alpha \leq c_k \leq 1 - \alpha$, which comes from the fact that, in case a2, $(c_k - \alpha)/(1 - \alpha) \geq 0$ and $c_k/(1 - \alpha) \leq 1$, respectively. Hence, $f_C(c_k) = 1/(1 - \alpha)$ and $f(\bar{c}|c_k) = (1 - \alpha)/\alpha$. Eq. (22) is finally given

by

$$\begin{aligned} & \int_{(c_k - \alpha)/(1 - \alpha)}^{c_k/(1 - \alpha)} \frac{(\alpha - c_k + \bar{c} - \alpha\bar{c})^{N-1} (1 - \alpha)}{\alpha^N} d\bar{c} \\ &= \frac{1}{N} \end{aligned} \quad (24)$$

and this gives the second line in Eq. (12) which holds for $\alpha \leq c_k \leq 1 - \alpha$, due to the condition on $f_C(c_k)$.

a3) For a given value of $\alpha \in (0, 0.5]$, \bar{c} now spans in the range $(c_k - \alpha)/(1 - \alpha) \leq \bar{c} \leq 1$ and the condition for c_k ($f_C(c_k)$) is $1 - \alpha < c_k \leq 1$. Hence, $f_C(c_k) = (1 - c_k)/(\alpha(1 - \alpha))$ and $f(\bar{c}|c_k) = (1 - \alpha)/(1 - c_k)$. Eq. (22) is obtained as

$$\begin{aligned} & \int_{(c_k - \alpha)/(1 - \alpha)}^1 \frac{(\alpha - c_k + \bar{c} - \alpha\bar{c})^{N-1} (1 - \alpha)}{(1 - c_k) \alpha^{N-1}} d\bar{c} \\ &= \frac{(1 - c_k)^{N-1}}{N \alpha^{N-1}} \end{aligned} \quad (25)$$

and this gives the third and last line in Eq. (12) which holds for $1 - \alpha < c_k \leq 1$, due to the condition on $f_C(c_k)$.

The derivation for case b is obtained through the same procedures by considering the diagram for cases b1, b2 and b3 in Fig. 6.