Alano: A Fast Neighbors Discover Algorithm In An Energy-Restricted Large-Scale Network

Abstract—**Neighbor discovery is a fundamental step in constructing wireless sensor networks and many algorithms have been proposed to minimize discovery latency. However, few of them can be applied to an energy-restricted large-scale network, which is more appealing and promising due to the development of intelligent devices. In an energy-restricted large-scale network, a node has limited power supply and can only discover nodes within its range; additionally, the discovery process may fail if surplus communication exists on the wireless channel. These factors make neighbor discovery a challenging task in establishing the networks.**

In this paper, we propose Alano, a nearly optimal algorithm for a large-scale network on the basis of nodes' distributions. When nodes have same energy constraints, we modify Alano by Relaxed Difference Set (RDS) (denote as RDS-Alano); while we present a Traversing Pointer (TP) based Alano (denote as TP-Alano) when the energy constraints are different. We compare Alano with the state-of-the-art protocols through extensive evaluations, and the results show that Alano achieves at least 31.35% lower discovery latency and it has higher performance regarding quality (discovery rate) and scalability.

I. Introduction

The popularity of Internet of Things (IoT) has turned people's attention back to wireless sensor networks [1], with a wide range of applications such as volcanic investigation [2], seismic detection [3], agriculture monitoring [4], etc.

Neighbor discovery is a crucial step by which a node discovers its neighboring nodes before processes like broadcasting and peer-to-peer communication, thus it is fundamental in constructing a wireless sensor network. In this paper, we study it in an energy-restricted large-scale scenario, where nodes are aware of energy consumption and are multi-hop connected.

Unfortunately, despite extensive research, neighbor discovery in a large-scale network remains an open problem. The existing algorithms can be classified into two categories: deterministic, and probabilistic. In the deterministic algorithms [5]–[9], sensors take actions based on deterministic sequences. However, most deterministic algorithms are designed only for two nodes and directly applied to multi-node scenario. Probabilistic algorithms handle neighbor discovery in a clique of n nodes [10]–[13] , i.e. every two nodes are neighbors, and utilize the global number n to compute an optimal probability for action decisions. However in a large-scale network, a node is only able to sense nodes within it's range. In addition, some existing algorithms [10], [11] do not consider energy consumption of the neighbor discovery process. In wireless sensor networks, sensors have limited energy and are recharged frequently. Attempting for neighbor discovery process consistently will be energy-consuming.

Therefore, we look into the existing neighbor discovery algorithms and find that the key issue lies in the way to deal with collisions in the large-scale networks (collisions result in CSMA [14] to wait for more time). This issue is due to three reasons. First, transmission signals fade with distance and simultaneous transmissions will cause collisions among various nodes. Deterministic algorithms aiming at two nodes [6], [9] fail to reduce such collisions. Some beacon-based algorithms [5], [7], [8] do not meet this issue but the time slot is 40 times larger and still result in high latency. Second, a large-scale network is not one-hop connected and a node can only discover its neighbors within it's range. Probabilistic algorithms [11]–[13] assuming a small clique network fail in estimating the number of neighbors, and thus can not reduce the collision effectively since the number of neighbors plays a vital role in how many collisions will occur at the same time. Third, nodes are powered by limited energy and they typically try to find neighbors only for a fraction of time. Energy consumption and neighbor discovery quality are a paradox in existing algorithms, since collisions cause great energy consumption and if taken energy consumption into account, neighbor discovery will become even ineffective.

We conduct experiments on the existing algorithms [5]–[9] and confirm this issue. By experiment, we find these algorithms reduce collisions either insufficiently or excessively, both resulting in a high latency. Collisions of Hedis [9] and PND [13] happen as frequently as 99.8% and 43.4% respectively. Hello [8] and Aloha-like [12] show a high idle rate (none of the neighbors are transmitting) as 94.6% and XX% of the time, which reduce the collisions excessively. That is because of failing to deal with collisions, existing work wastes time and energy, and cannot achieve low latency and energy efficiency for large-scale networks at the same time.

Our insight idea is that, by estimating the expected number of neighbors of a node and synchronizing the time it turns on the radio with its neighbors, we can achieve both low-latency and energy-efficiency for neighbor discovery. We first take the distributions of nodes into consideration. As studied in [15], nodes in a wireless sensor network are likely to follow a uniform or a Gaussian distribution for detecting aims. According to the local density, a node can estimate the number of neighbors and calculate an optimal probability for action decisions. Then based on this, we propose Alano, a nearly optimal probability based algorithm for a large-scale network. Finally, we involve the duty cycle mechanism, i.e. the fraction of time the radio is on (also called a sensor wakes up), and modify Alano by deterministic techniques to synchronize the wake-up time between neighbors. Specifically, if all nodes have the same (symmetric) duty cycle, such as a batch of sensors have a default duty cycle setting, we propose a Relaxed Difference Set based algorithm (called RDS-Alano); if nodes have different (asymmetric) duty cycles, such as a sensor adjusts the duty cycle by the remaining energy, we propose a Traversing Pointer based algorithm (called TP-Alano).

In our simulations we have found Alano has 31.35% to 85.25% lower latency, higher discovery rate, and better scalability in large scale networks. In comparison to the state-of-the-art algorithms [7]–[9], [12] Alano reaches nearly 100% discovery rate within half time. When the number of nodes increases from 1000 to 9000, Alano shows 4.68 times to 6.51 times lower latency for neighbor discovery.

The contributions of the paper are summarized as follows:

1) We utilize distribution of nodes and propose Alano, a nearly optimal algorithm that achieves low-latency neighbor discovery for a large-scale network;
2) In an energy-restricted large-scale network, we propose RDS-Alano for symmetric nodes and TP-Alano for asymmetric nodes. Both algorithms achieve low latency for discovering neighbors and can prolong node's lifetime;
3) We conduct experiments for fundamental observation and extensive simulations for large-scale networks. Alano achieves lower latency, higher discovery rate, and better scalability, which promises a potential scalability of IoT in the future work.

The remainder of the paper is organized as follows. The coming section highlights some related works and puts forward vital problems that the existing algorithms remain. The system model and basic definitions are introduced in Section III. We present Alano and show the method to combine nodes' distribution in Section IV. We propose two modified algorithms (RDS-Alano, TP-Alano) for a energy-restricted large-scale network for both symmetric and asymmetric nodes in Section V. The extensive simulation results are shown in Section VI and we conclude the paper in Section VII.

II. Related Work

Existing neighbor discovery algorithms can be classified into two categories, deterministic algorithms and probabilistic algorithms.

Deterministic algorithms adopt certain mathematic tools to ensure discovery between every two neighbors. The first tool is called quorum system [16], [17] : for any two intersected quorums, two neighboring nodes could choose any quorum in the system to design the discovery schedule. Hedis [9] is a typical one. Another important tool is co-primality where two co-prime numbers are chosen by the neighbors to design the discovery schedule, and they can discover each other within a bounded latency by the Chinese Remainder Theorem [18]. Some representative algorithms are Disco [5], U-Connect [6], and Todis [9]. These algorithms hold an obvious advantage that they can guarantee fast discovery for two nodes within a bounded latency.

However, there are some weaknesses in these algorithms when applied to large-scale networks. Hedis chen2015heterogeneous only assumes when two nodes turn on the radio at the same time and they find each other. But in reality neighbors play asymmetric roles that one is transmitting and the other is receiving. U-Connect [6] consider the transmitting and receiving roles. However different from two-node scenario, collisions will happen when many nodes are transmitting simultaneously. Some deterministic algorithms propose a beacon-based transmission protocol. For example, Disco [5] assumes that a node has a capability to send a beacon (one or a few bits) at both beginning and end of an active time slot, and the assumption is adopted in SearchLight [7], and Hello [8]. Under this assumption, the listening time slot is much larger than the transmitting time slot, which makes a complete slot 40 times larger than that of U-Connect [6] and thus results in a high latency.

Another category is probabilistic algorithms [10]–[13] which utilize probability techniques to promote the randomness of discovering the neighbors. Birthday protocol [10] is one of the earliest algorithms that works on the birthday paradox, i.e. the probability that two people have the same birthday exceeds $\frac{1}{2}$ among 23 people. Following that, smarter probabilistic algorithms are proposed, such as Aloha-like [11], [12], PND [13]. Particularly, Aloha-like [11] does not consider the energy consumption, which is later extended to an energy-restricted network by [12].

The difference from deterministic algorithms is that, probabilistic algorithms show an significant strength for a network consisting of multiple nodes. However, probabilistic algorithms only present an expectation discovery latency and they can not guarantee a good latency bound. In addition, most of the existing algorithms assume the network is a clique, which implies any two nodes are neighbors. This assumption can hardly depict a large-scale network due to the limited communication range of a node. Some work adopts the received signal strength to decide how far a node can transmit [19], and in the protocol model, it is simplified that two nodes can communicate if their distance is no larger than a threshold.

III. Preliminaries

In this section, we first describe the system model of an energy-restricted large-scale network. Then we formulate the Neighbor Discovery problem formally. The notations are listed in Table I.

A. System Model

We introduce three important factors in an energy-restricted large-scale network.

Communication: When multiple nodes communicate simultaneously, transmission may fail due to the communicational interference. We adopt the protocol model (also called graph model, unit-disk model [20], [21]) to describe the process, which assumes a node u_i can receive u_j's message

TABLE I
NOTATIONS FOR NEIGHBOR DISCOVERY

Notation	Description
N	The number of nodes in the network
u_i	Node u_i with ID i
$\widetilde{n_i}$	Node u_i's expected number of neighbors
S_i	The set of u_i's neighbors
t_0	The length of a time slot
δ_{ij}	The transmission time drift between u_i and u_j
t_i^s	The start time of node u_i
$L(i,j)$	The discovery latency that node u_i discovers node u_j
$L(i)$	The discovery latency that node u_i discovers all neighbors
θ	The pre-defined global duty cycle
θ_i	Node u_i's local duty cycle
W	The time slot spent by a node discovering all neighbors
M	Neighboring matrix, $M_{ij}=1$ means u_i and u_j are neighbors

successfully if u_j is the only transmitter that is within u_i's communication range. The protocol model is a popular one that enables the development of efficient algorithms for crucial networking problems. Some other models, such as the signal to interference plus noise ratio (*SINR*) model, are more complicated and lack good algorithmic features. In addition, it is shown that these models can be transformed to the protocol model by particular means in [22].

Network connectivity: A node can only discover nodes that are within its range under the protocol model. In a large-scale network, two nodes may not be connected directly and we call two nodes connected by one-hop communication as *neighbors*. Thus, a large-scale network is always *partially-connected*, contrary to a fully-connected network in which any two nodes are neighbors.

Energy-restricted: A node in the network has limited energy, and can turn on or off its radio to save energy [23]. When a node turns on the radio, it can transmit a message including its identifier (ID), or listen on the channel to receive a neighbor's message.

Technically speaking, we assume an energy-restricted large-scale network consisting of N nodes as set $U = \{u_1, u_2, \dots, u_N\}$. Nodes are distributed in a large area and they can communicate through a fixed wireless channel. We assume the locations of the nodes obey some distributions, such as uniform distributions and Gaussian distribution [15]. Suppose a node has a fixed communication range Δ and two nodes u_i, u_j are neighbors if their distance suits $d(u_i, u_j) \leq \Delta$. We can use a symmetric matrix $M_{N\times N}$ to represent nodes' neighboring relations as:

$$M_{i,j} = \begin{cases} 1 & u_i\ and\ u_j\ are\ neighbors \\ 0 & otherwise \end{cases}$$

Suppose time is divided into slots of equal length t_0 [24], which is sufficient to finish a complete communication process (one node transmits a message including its ID and a neighbor receives the message). A node who turns its radio on can choose to be transmitting state or listening state:

- **Transmitting state:** a node transmits (broadcasts) a package containing its ID on the channel;
- **Listening state:** a node listens on the channel to receive message from neighbors.

In the protocol model, a node u_i can discover its neighbor u_j in time slot t if and only if u_j is the only neighbor of u_i that transmits and u_i listens in the slot.

A node has limited energy and it has to turn off the radio to save energy for most of the time. We assume a *duty schedule* for a node u_i is a pre-define sequence $S_i = \{s_i(t)\}_{0\leq t<T}$ of period T, in which

$$s_i^t = \begin{cases} S & u_i\ turns\ off\ the\ radio\ in\ slot\ t \\ T & u_i\ is\ in\ transmitting\ state\ in\ slot\ t \\ L & u_i\ is\ in\ listening\ state\ in\ slot\ t \end{cases}$$

duty cycle is defined as the fraction of time a node turns its radio on, which is formulated as:

$$\theta_i = \frac{|\{t : 0 \leq t < T, s_i(t) \in \{T, L\}\}|}{T}.$$

If all nodes have the same duty cycle all the time, i.e. $\theta_i = \theta_j$ for any nodes u_i, u_j, we call them *symmetric nodes*. Otherwise, they are *asymmetric nodes*. Denote the start time of node u_i as t_i^s and denote δ_{ij} as the time drift between a pair of neighbors u_i, u_j, i.e. $\delta_{ij} = t_i^s - t_j^s$.

B. Problem Definition

A node u_i executes operations (turning off radio, transmitting, or listening) according to the pre-defined duty schedule S_i. When u_i starts the neighbor discovery process, denote $L(i,j)$ as the slot cost to find the neighbor u_j and we define **discovery latency** of node u_i as the time to discover all neighbors:

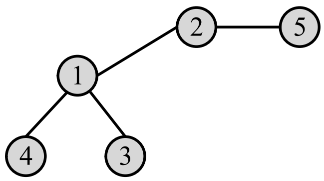
$$L(i) = \max_{j:M_{i,j}=1} L(i,j).$$

It is to be noted that, neighbor discovery is not bidirectional and any pair of neighbors have to discover each other separately. We formulate the neighbor discovery problem for node u_i as:

Problem 1: For a node u_i and the set of its neighbors $S_i = \{u_j | d(u_i, u_j) \leq \Delta\}$, design duty schedules for all nodes such that: $\forall u_j \in S_i$:

$$\exists t\ s.t.: \\ s_i(t) = L, s_j(t) = T, and\ \forall u_k \in S_i, u_k \neq j : s_k(t) \in \{L, S\}.$$

Fig.1 shows an example of 5 nodes, the topology is depicted in Fig.1(a), and Fig.1(b) describes the neighbor discovery process. The duty cycle is set as 0.25 and nodes start in different time slots. The designed duty schedule for node u_1 is $S_1 = \{T, S, S, S, S, S, T, S, S, T, S, S, S, T, ...\}$. Suppose time drift between node u_1 and u_4 is $\delta_{14} = 1$; u_1 u_2 start in the same time slot. In slot 12, node u_5 discovers neighbor u_2, but node u_1 cannot discover u_2 since another neighbor u_3 is in transmitting state simultaneously.



(a) The topology of a simple wireless network

Time	…	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	…
Node 1			T	S	S	S	S	S	L	S	S	L	S	S	S	T	…
Node 2			S	S	L	S	S	S	T	S	L	S	S	S	S	L	…
Node 3	…	S	S	S	T	S	S	S	T	S	L	S	S	S	L	S	…
Node 4				S	S	S	S	S	S	S	S	S	T	T	T	T	…
Node 5							S	S	L	S	S	S	T	S	S	S	…

(b) Neighbor discovery process

Fig. 1. An example of neighbor discovering process. S, T and L represents Sleep pattern, Transmitting state and Listening state in wake-up pattern respectively.

IV. Alano Algorithm for a large-scale Network

In a large-scale network, nodes are not fully-connected and communications may fail due to concurrent transmissions. When we do not consider the energy constraints of nodes, we propose a nearly optimal algorithm for a large-scale network, which implies $\theta_i = 1$ for all node u_i. Suppose the locations of nodes obey some distribution, we propose Alano algorithm and analyze its performance for two common used distribution (uniform distribution and Gaussian distribution).

A. Alano Algorithm

Suppose the locations of nodes obey some distribution and each node u_i is aware of its position coordinates (x_i, y_i). Then u_i could compute its local density by the following general function:

$$f(x,y) = \begin{cases} \varphi(x,y) & (x,y) \in D \\ 0 & (x,y) \notin D \end{cases}$$

where (x, y) is a position coordinate, and D is the network covering area. Denote the range of u_i's neighbors' positions as R_i, and any neighbor with coordinates $(x, y) \in R_i$ suits:

$$(x - x_i)^2 + (y - y_i)^2 \leq \Delta^2$$

where Δ is the communication range. Then, node u_i's expected number of neighbors (denote as $\widetilde{n_i}$) is:

$$\widetilde{n_i} = N \iint_{R_i} f(x,y)\,dx\,dy - 1.$$

In a large-scale network, we ignore the boundary area of the network. Note that, when the network covering area D is much larger than the area R_i of node u_i, we have:

$$\widetilde{n_i} \simeq N\pi\Delta^2\varphi(x_i, y_i). \tag{1}$$

We present **Alano**, a randomized algorithm for node u_i in Alg. 1. By computing the expected number of nodes, u_i turn to be in transmitting state or listening state according to the generated probability in Line 1.

Algorithm 1 Alano Algorithm

1: Set tranmist probability $p_t^i := \frac{1}{\widetilde{n_i}+1}$, $t := 0$;
2: **while** $t \leq T$ **do**
3: Generate a random number $\epsilon \in (0, 1)$;
4: **if** $\epsilon < p_t$ **then**
5: Transmit a message containing u_i's information including its ID through the channel;
6: **else**
7: Listen on the channel. If receive a message successfully, decode the message and record the sender's ID;
8: **end if**
9: $t := t + 1$;
10: **end while**

In the following parts, we first consider a general situation that nodes in the network are uniform distributed. We derive a proof that the probability chosen in Alano is the optimal one and show the discovery latency will not be much larger than its expectation. T in Alg. 1 Line 2 is the time threshold and can be set as the latency bound. Then we analyze a more common situation that nodes obey Gaussian Distribution, we present an approximation analysis that the discovery latency is not be much larger than that of uniform distribution.

B. Analysis for Uniform Distribution

Uniform distribution is a basic one for deployment of wireless networks. For instance, to monitor an unknown area, many sensors are deployed uniformly to collect information, such as temperature and humidity [25]. The nodes are evenly deployed and the density function is:

$$f(x) = \begin{cases} \frac{1}{A} & (x,y) \in D \\ 0 & (x,y) \notin D \end{cases}$$

where A is the area of D. By Eqn. (1), u_i's expected number of neighbors is $\widetilde{n_i} = \frac{N\pi\Delta^2}{A}$ and the probability in Line 2 is set as $p_t^i = \frac{1}{\widetilde{n_i}+1} = \frac{A}{N\pi\Delta^2+A}$.

Lemma 1: Alg. 1 achieves optimal discovery latency for uniform distribution by setting $p_t^i = \frac{1}{\widetilde{n_i}}$.

Proof: For any two nodes u_i abd u_j in the uniform distribution, we get $\widetilde{n_i} = \widetilde{n_j} = \widetilde{n}$ and $p_t^j = p_t^i = p_t = \frac{1}{\widetilde{n}+1}$.

From Alg. 1, the probability that node u_i discovers a specific neighbor (such as u_j) successfully in a time slot (denote as p_s) is:

$$p_s = p_t(1 - p_t)^{\widetilde{n}-1}(1 - p_t).$$

In order to compute the maximum probability to discover a node, we compute the differential function of p_s as:

$$\frac{d(p_s)}{d(p_t)} = (1 - p_t)^{\widetilde{n}} - \widetilde{n}p_t(1 - p_t)^{\widetilde{n}-1}.$$

It is easy to verify that when $p_t = \frac{1}{\widetilde{n}+1}$, p_s gets the maximum value:

$$p_s = \frac{1}{\widetilde{n}+1}(1-\frac{1}{\widetilde{n}+1})^{\widetilde{n}} \approx \frac{1}{(\widetilde{n}+1)e}.$$

Therefore, the probability chosen in Alano algorithm for transmitting is optimal. ■

Theorem 1: Alg. 1 discovers all neighbors for node u_i within $T = \Theta(n \ln n)$ time slots with high probability.

Proof: First, we show that the expected discovery latency of a node u_i is bounded by $O(n \ln n)$.

Let W be a random variable that denotes the time a node spends discovering all neighbors. Denote W_j as a random variable representing the cost number of the time slots to discover a new neighbor after $j-1$ neighbors have been discovered. It is easy to check that W_j follows Geometric distribution with parameter $p(j)$ where $p(j) = (\widetilde{n}-j+1)p_s$ [26]. The expectation of W_j can be computed as:

$$E[W_j] = \frac{1}{p(j)} = \frac{1}{(\widetilde{n}-j+1)p_s}.$$

Then, the expectation discovery latency of node u_i is:

$$E[W] = \sum_{j=1}^{\widetilde{n}} E[W_j] \approx (\widetilde{n}+1)e(\ln(\widetilde{n}+1)+\Theta(1)) = \Theta(n \ln n).$$

Thus the expected discovery latency is bounded by $O(n \ln n)$.

Then, we show that node u_i can discovery all neighbors bounded by $O(n \ln n)$ time slots.

Since W_j follows Geometric distribution, $Var[W_j] = \frac{1-p_j}{p_j^2}$, and the variance of discovery latency of W is computed as:

$$Var[W] = \sum_{j=1}^{n} Var[W_j] \leq \frac{\pi^2}{6p_s^2} - \frac{H_n}{p_s}.$$

By *Chebyshev's inequality*, the probability that the discovery latency is 2 times larger than the expectation is

$$P[W \geq 2E[W]] \leq \frac{Var[W]}{E[W]^2} \leq \frac{\pi^2}{6H_n^2} - \frac{p_s}{H_n}.$$

where H_n is the n-th Harmonic number, i.e., $H_n = lnn + \Theta(1)$. For a large n, $P[W \geq 2E[W]]$ is close to 0. That is, the time for a node to discover all neighbors is very likely to be smaller than 2 times of the expectation. Therefore, W is also bounded by $O(n \ln n)$ with high probability. ■

C. Analysis of Gaussian Distribution

Gaussian distribution is common adopted in wireless network. For instance, an intrusion detection application needs larger detection probability around important entities [15]. We assume the positions of nodes obey 2D Gaussian distribution, and we present a theoretical proof that the discovery latency is not much larger than that of uniform distribution. Without loss of generality, we consider $(x, y) \sim N(0, 1, 0, 1, 0)$.

Theorem 2: Alg. 1 discovers all neighbors for node u_i within $T = \Theta(n \ln n)$ time slots with high probability.

Proof: Denote the approximate neighbors of node u_i as set $S_i = \{u_j | d(u_i, u_j) \leq \Delta\}$. When nodes obey Gaussian distribution, the probability that node u_i discovers a certain neighbor node u_j successfully in a time slot (denote as p_s) can be formulated as:

$$p_s^i = (1-p_t^i) \cdot p_t^j \cdot \prod_{u_j \in S_i, u_j \neq u_k} (1-p_t^k).$$

Denote $p_t^{max} = \max_{u_j \in S_i}\{p_t^j\}, p_t^{min} = \min_{u_j \in S_i}\{p_t^j\}$, for every $u_j \in S_i$, we have:

$$(1-p_t^i)p_t^{min}(1-p_t^{max})^{\widetilde{n_i}-1} \leq p_s^i$$
$$p_s^i \leq (1-p_t^i)p_t^{max}(1-p_t^{min})^{\widetilde{n_i}-1}$$

Denote:

$$P = (1-p_t^i)p_t^{min}(1-p_t^{max})^{\widetilde{n_i}-1}.$$
$$Q = (1-p_t^i)p_t^{max}(1-p_t^{min})^{\widetilde{n_i}-1}.$$

We derive the expectation of W_j for $1 \leq j \leq n_i$ as:

$$\frac{1}{(\widetilde{n_i}-j+1)Q} \quad \leq \quad E[W_j] \quad \leq \quad \frac{1}{(\widetilde{n_i}-j+1)P}.$$

Combine the equations to derive:

$$\frac{1}{Q}H_n \quad \leq \quad E[\sum_{j=1}^{\widetilde{n_i}} W_j] \quad \leq \quad \frac{1}{P}H_n.$$

Since:

$$p_t^{min} = \frac{1}{\widetilde{n}_{max}+1}, \quad p_t^{max} = \frac{1}{\widetilde{n}_{min}+1}.$$

where $\widetilde{n}_{max} = \max\{\widetilde{n}_j | u_j \in S_i\}$, $\widetilde{n}_{min} = \min\{\widetilde{n}_j | u_j \in S_i\}$. Thus:

$$\widetilde{n_i} \simeq N\pi\Delta^2 \frac{1}{2\pi} \mathrm{e}^{-\frac{x_i^2+y_i^2}{2}}.$$
$$\widetilde{n}_{max} \simeq N\pi\Delta^2 \frac{1}{2\pi} \mathrm{e}^{-\frac{(\sqrt{x_i^2+y_i^2}-\Delta^2)^2}{2}} = \alpha\widetilde{n_i}.$$
$$\widetilde{n}_{min} \simeq N\pi\Delta^2 \frac{1}{2\pi} \mathrm{e}^{-\frac{(\sqrt{x_i^2+y_i^2}+\Delta^2)^2}{2}} = \beta\widetilde{n_i}.$$

and we get:

$$\frac{1}{Q}H_n \simeq \beta \mathrm{e}^{\frac{1}{\alpha}} n \ln n, \quad \frac{1}{P}H_n \simeq \alpha \mathrm{e}^{\frac{1}{\beta}} n \ln n.$$

Thus $E[W]$ can be bounded within $O(n \ln n)$ time slots with high probability. Similarly the bounded latency can be proved to be $W = O(n \ln n)$ by *Chebyshev's inequality* in the same way as uniform distribution. ■

V. Modified Alano for An Energy-Restricted Network

In an energy-restricted network, nodes have limited energy and designing duty schedule for a node needs to take its duty cycle into account. Obviously, a lower duty cycle implies a larger discovery latency since the node turns its radio off for more time during the schedule.

In the preceding section, energy constraint is not a crucial factor in Alano; and we are to modify Alano for both symmetric nodes and asymmetric nodes. Our initiative idea is to align the slots that the radio is on with the neighbors in a bounded time; then invoke Alano algorithm to achieve low-latency neighbor discovery.

A. RDS-Alano for Symmetric Nodes

Symmetric nodes have the same duty cycle $\theta_i = \theta_j = \theta$, $\forall u_i, u_j$. We utilize Relax Difference Set (RDS) to align time slots that nodes are in transmitting or listening state.

RDS is an efficient tool to construct cyclic quorum systems [16], [17]. The definition is:

Definition 1: A set $R = \{a_1, a_2, ..., a_k\} \subseteq Z_n$ (the set of all nonnegative integers less than n) is called a RDS if for every $d \neq 0$ (mod n), there exists at least one ordered pair (a_i, a_j) such that $a_i - a_j \equiv d$ (mod n), where $a_i, a_j \in D$.

It has been proved that any RDS must have cardinality $|R| \geq \sqrt{N}$ [17]. We present a linear algorithm to construct a RDS with cardinality $\lceil \frac{3\sqrt{N}}{2} \rceil$ under Z_N in Alg. 2.

Algorithm 2 RDS construction under Z_N

1: $R := \emptyset$;
2: $\lambda := \lceil \sqrt{N} \rceil$,$\mu := \lceil \frac{\lceil \sqrt{N} \rceil}{2} \rceil$;
3: **for** $i = 1 : \lambda$ **do**
4: $\quad R := R \cup i$;
5: **end for**
6: **for** $j = 1 : \mu$ **do**
7: $\quad R := R \cup (1 + j * \lambda)$;
8: **end for**

The intuitive idea of Alg. 2 can be described as Fig. 2. The framed elements are selected as Line. 4 and Line. 7. We show the correctness of the construction formally.

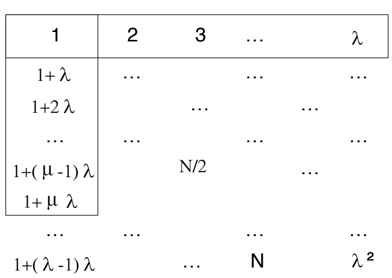


Fig. 2. An Sketch of RDS construction in Alg. 2

Lemma 2: Set $R = \{r_0, r_1, ..., r_{\lambda+\mu-1}\}$ constructed in Alg. 2 is a RDS, where $|R| = \lambda + \mu = \lceil \sqrt{N} \rceil + \lceil \frac{\lceil \sqrt{N} \rceil}{2} \rceil \approx \lceil \frac{3\sqrt{N}}{2} \rceil$.

Proof: Obviously, if there exists one ordered pair (a_i, a_j) satisfying $a_i - a_j \equiv d$ (mod N), an opposite pair (a_j, a_i) exists such that $a_j - a_i \equiv (N - d)$ (mod N). Thus we only need to find at least one ordered pair (a_i, a_j) for each $d \in [1, \lfloor N/2 \rfloor]$.

In the construction, λ in Line 2 is the smallest integer satisfying $\lambda^2 \geq N$. Every d within range $[1, \lfloor N/2 \rfloor]$ can be represented as: $d = 1 + j \times \lambda - i$, where $1 \leq j \leq \mu, 1 \leq i \leq \lambda$. Thus, there exists $a_j = 1 + j \times \lambda$ from Line. 4 and $a_i = i$ from Line. 7 satisfying $a_j - a_i \equiv d$. Then, the lemma can be derived. ■

For symmetric nodes with duty cycle θ, we present a RDS based Alano (RDS-Alano) algorithm as Alg. 3.

In Alg. 3, RDS is used to construct a deterministic schedule for a node to turn on its radio in every period of length T, and Alano is utilized as a probabilistic strategy to determine whether it is in transmitting state or listening state.

Algorithm 3 RDS Based Alano Algorithm

$T := \lceil \frac{9}{4\theta^2} \rceil$;
Invoke Alg. 2 to construct the RDS $R = r_0, r_1, ..., r_{\lceil \frac{3\sqrt{T}}{2} \rceil}$ under Z_T;
$t := 0$;
while $True$ **do**
$\quad$**if** $(t + 1) \in R$ **then**
$\quad\quad$Invoke Alg. 1 to determine transmission state;
$\quad$**else**
$\quad\quad$Sleep;
$\quad$**end if**
$\quad$t := (t + 1) % T;
end while

We derive discovery latency bound for RDS-Alano:

Theorem 3: RDS-Alano guarantees that the discovery latency of a node is bounded within $O(\frac{n \log n}{\theta^2})$ with high probability.

Proof: First, we verity that the duty cycle in RDS-Alano (denote as $\widetilde{\theta}$) is

$$\widetilde{\theta} = \frac{|RDS|}{|T|} = \frac{\lceil \frac{3\sqrt{T}}{2} \rceil}{T} = \theta.$$

For any pair of neighbor nodes (u_i, u_j), we can find an ordered pair (r_i, r_j) from their respective RDS such that $r_i - r_j \equiv \delta_{ij}$ (mode T), where δ_{ij} is the time drift. This implies any neighbor nodes can turn on their radios in the same time slot for at least once during every period of length T. Regarding each period of T time slots as a 'super' slot of Alano algorithm, we can derive that the discovery latency is bounded within $O(\frac{n \ln n}{\theta^2})$ slots with high probability by combining the analysis of Alano. ■

Remark 1: In a RDS, a node can discovery its neighbors in different time slots. When treating a period of T as a 'super' slot of Alano, there may be less than the total neighbors in each wake-up sub-slot, resulting less collisions and lower latency compared to when all the neighbors wake up in the same sub-slot. Thus the latency bound can not be larger.

B. TP-Alano for Asymmetric Nodes

Considering a more practical network where nodes can adjust their duty cycles, we present a traversing pointer method to align time slots that nodes are in transmitting or listening state for asymmetric nodes.

For a more practical scenario, the nodes in a wireless sensor networks for instance, are assigned to diverse tasks such as temperature measurement, sunshine collection, etc., and thus ought to have asymmetric capability of battery-management with local duty cycle θ_i.

Suppose the duty cycle of node u_i is θ_i, we present a traversing pointer based Alano (TP-Alano) algorithm as Alg. 4. In each period of T slots, a node turns on its radio in two different time slots, one of which is the first slot of each period and the other one is a traversing slot that changes from period to period (as described in Line. 5).

Algorithm 4 Traversing Pointer Based Alano Algorithm

```
1: T :=Find the smallest prime ≥ 2/θ_i;
2: t := 0;
3: while True do
4:    t_1 := t%T;
5:    t_2 := ⌊t/T⌋%(T − 1) + 1;
6:    if t_1 = 0||t_1 = t_2 then
7:       Invoke Alg. 1 to determine transmission state;
8:    else
9:       Sleep;
10:   end if
11:   t := t + 1;
12: end while
```

We call the first time slot of each period as *fixed pointer* and the traversing slot as *traversing point*. The pointers are designed to guarantee that nodes u_i, u_j could turn on their radio simultaneously in very period of length T_iT_j. A sketch of the pointers is described in Fig. 3.

Note that, a period of T slots is constructed as Line 1 where we try to *find the smallest prime* $\geq \frac{2}{\theta_i}$, then it is likely to make the duty cycle of each period smaller than the expected one. This can be easily solved by selecting some random slots to turn on the radio for listening in each period T to conform to expected duty cycle.

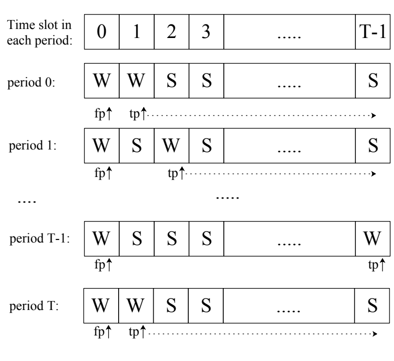


Fig. 3. A sketch of TP construction in Alg. 4

We show the discovery latency of TP-Alano algorithm as:

Theorem 4: TP-Alano guarantees that the discovery latency $L(i,j)$ is bounded within $O(\frac{nlogn}{\theta_i\theta_j})$ with high probability., where θ_i and θ_j are the duty cycles of a pair of neighbors (u_i, u_j) respectively.

Proof: We first prove that any pair of nodes (u_i, u_j) turn on their radios (for transmitting or listening) simultaneously for at least once in every period of length T_iT_j.

Case 1: $T_i \neq T_j$. Since T_i and T_j are different primes, according to Chinese Remainder Theorem [18], there exists a time slot $t_\tau \in [0, T_iT_j)$ satisfying:

$$0 = t_\tau \quad mod \quad T_i. \tag{2}$$

$$\delta_{ij} = t_\tau \quad mod \quad T_j. \tag{3}$$

Thus, there exists a fixed pointer of node u_i and a fixed pointer of node u_j in which both nodes turn on the radios in every period of length T_iT_j .

Case 2: $T_i = T_j$. Since $T_i = T_j = T$, if the time drift between u_i and u_j is $\delta_{ij} = 0$, the fixed pointers of u_i and u_j will be the same in every period of length T. Otherwise, since the traversing point will traverse all the time slots once during period of length $(T-1)T$, there exists a traversing point of u_i and a fixed pointer of u_j satisfying that both nodes turn on the radios simultaneously in every period of length $(T-1)T$; similarly a traversing point of u_j and a fixed pointer of i satisfying that they both turn on the radios.

Thus for any pair of neighbor nodes (u_i, u_j), they turn on their radios for transmitting or listening for at least once in every period of length T_iT_j. Regarding the whole period T_iT_j as a 'super' slot of Alano, we derive that the discovery latency is bounded within $O(\frac{n \ln n}{\theta_i\theta_j})$ with high probability. ■

VI. Evaluation

In this section, we first conduct experiments and confirm our basic analysis. Then we evaluate the algorithms in large-scale networks by simulations.

A. Experiments for Fundamental Observation

In our experiment, we implemented 16 EZ240 sensors, each of which uses 8MHz MSP430F1611 as the microprocessor and is equipped with 10K RAM, 48K ROM and 1 M Flash. CC2420 is used as the communication module with the IEEE 802.15.4 protocol. RTIMER_ARCH_CECOND is the clock frequency and a time slot is set as 500/RTIMER_ARCH_CECOND, which is around 0.5ms in the real world. As discussed in section II, the time slot of beacon-based algorithms, e.g. searchlight [7], is set as 20ms. We can see from the APIs in MAC layer that CSMA mechanism makes the process to wait a random time when a collision happens, which implies collisions result in a high latency.

Algorithms	**Alano**	**Hedis**	**Searchlight**	**ALOHA**
Average latency	1.278	2.447	8.58	10.07
Maximum latency	1.52	8.36	10.11	16.13
Minimum latency	1.10	0.73	6.31	5.11

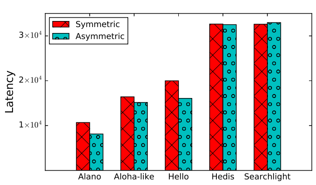
We compared Alano with Searchlight [7], Hedis [9] and ALOHA [12] in the experiment. We can see the result from the Table VI-A that Alano outperforms the other algorithms. Hedis shows low latency as well since it has an ideal latency bound for two nodes, and in a small-scale network its weakness of handling collisions is not obvious. Searchlight is a typical beacon-based algorithm with larger time slot, which results in high latency. ALOHA is designed for a clique network and the probability adopted is not optimal.

B. Simulations for Large-scale Networks

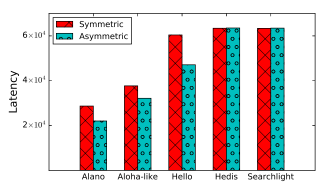
To simulate a large-scale network, we implemented Alano in C++ and evaluated the algorithms in a cluster of 9 servers, each equipped with an Intel Xeon 2.6GHz CPU with 24 hyper-threading cores, 64GB memory and 1T SSD.

We simulated a network that follows uniform distribution and Gaussian distribution respectively. For uniform distribution, we suppose 500 nodes are distributed in an area of $100m \times 100m$ and each node's communication range is $\Delta = 10m$. For Gaussian distribution, we suppose 1000 nodes are distributed in the same area, but each node's communication range is $\Delta = 5m$. We set the Gaussian distribution as $N(50, 15^2)$ in our evaluation. We set the duty cycle of a node to be 0.1 for symmetric nodes; for asymmetric nodes, we set their duty cycle randomly from 0.05 to 0.15 with step 0.02. These settings make the network more complicated and realistic than that in [6]–[13], [27].

We evaluated discovery latency of Alano, Aloha-like [12], Hello [8], Hedis [9], and Searchlight [7] in the generated networks. Hello and Searchlight has a beacon transmission of $0.54ms$ at the beginning and end of each slot, and a beacon makes up about $1/40$ of a slot, so we divide each $\{ON\}$ slot into 40 mini-slots, and the node transmits at the first and last mini-slot, and listens in other mini-slots. In the following parts, we show that Alano has lower latency, higher discovery rate, and better scalability.



(a) Uniform Distribution

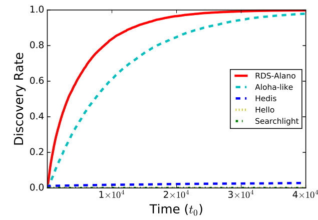


(b) Gaussian Distribution

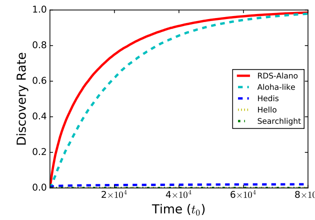
Fig. 4. Alano achieves lower latency.

1) Speed-Discovery Latency: When nodes follow uniform distribution, we show the discovery latency comparison for both symmetric nodes and asymmetric nodes in Fig. 4(a). From the figure, Alano achieves 54.64% to 1.95 times lower discovery latency for symmetric nodes, and 85.25% to 2.91 times lower discovery latency for asymmetric nodes. When nodes follow Gaussian distribution, as depicted in Fig. 4(b), Alano achieves 31.35% to 1.21 times lower discovery latency for symmetric nodes, and 45.94% to 1.88 times lower discovery latency for asymmetric nodes. The deterministic algorithms, Hello, Hedis and Searchlight, have high latency due to either collisions or larger time slots.

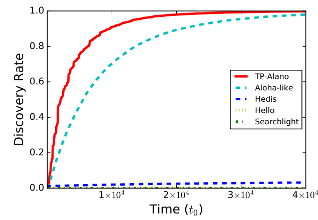
2) Quality-Discovery Rate: We use discovery rate to evaluate Alano's quality. Discovery rate of a node u_i is defined as the percentage of discovered neighbors over u_i's all neighbors. In Fig. 5, we increase the number of nodes from 500 to 1000 for the uniform distribution, and increase the number of nodes from 1000 to 2000 for the Gaussian distribution, the results show that Alano has higher discovery rate during the whole process for both uniform and Gaussian distributions.



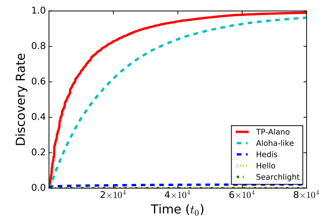
(a) Symmetric Nodes in Uniform Distribution



(b) Symmetric Nodes in Gaussian Distribution



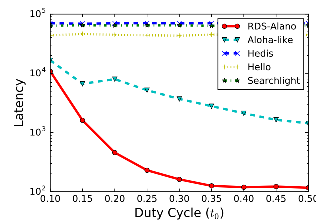
(c) Asymmetric Nodes in Uniform Distribution



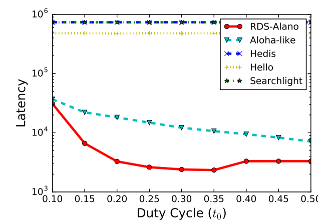
(d) Asymmetric Nodes in Gaussian Distribution

Fig. 5. Alano achieves higher discovery rate in larger networks.

3) Scalability-Duty Cycle and Network Density: We evaluated Alano's scalability regarding duty cycle and network density.



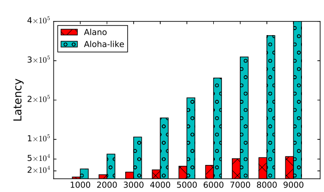
(a) Uniform Distribution



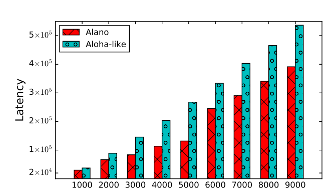
(b) Gaussian Distribution

Fig. 6. Alano achieves lower latency in different duty cycle.

Duty Cycle. When symmetric nodes with different duty cycles, Fig. 6 shows that Alano has lower latency. Compared with Aloha, Alano has from 53.66% to 11.23 times lower latency. The latency of Alano and Aloha generally decreases as the duty cycle increases, while Hello, Hedis and Searchlight have high latency due to the collision. In Gaussian distribution, Alano has a small twist with duty cycle 0.35, because when the duty cycle increases, nodes are more likely to transmit and therefore collide.



(a) Uniform Distribution



(b) Gaussian Distribution

Fig. 7. Alano achieves lower latency with different number of nodes.

Network Density. When the number of nodes increases, the network becomes denser. We choose Aloha-like algorithm for comparison because Hello, Hedis and Searchlight already have higher latency than Aloha-like when there are 500 nodes in uniform distribution and 1000 nodes in Gaussian distribution. As shown in Fig. 7(a), Alano achieves 4.68 times to 6.51 times lower discovery latency than Aloha-like algorithm for uniform distribution, when the number of nodes increases from 1000 to 9000. When the number of nodes increases from 1000 to 9000 for Gaussian distribution, Alano achieves 25.23% to 1.03 times lower discovery latency as shown in Fig. 7(b).

VII. Conclusion

In this paper, we initiate the study of neighbor discovery in an energy-restricted large-scale network. To begin with, we propose Alano for a large-scale network where nodes' distribution are utilized to trigger the design of a node's transmitting probability. For different distributions, such as uniform distribution and normal distribution, we show that Alano achieves nearly optimal discovery latency. Then, we propose two modification methods for a energy-restricted network on the basis of different duty cycle mechanisms: Relaxed Different Set based Alano (RDS-Alano) for symmetric nodes and Traversing Pointer based Alano (TP-Alano) for asymmetric nodes. We conduct extensive simulations to compare Alano with the state-of-the-art algorithms, the results show that Alano achieves better performance regarding discovery latency, discovery rate (quality), and scalability.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer networks*, vol. 38, no. 4, pp. 393–422, 2002.

[2] G. Werner-Allen, K. Lorincz, M. Ruiz, O. Marcillo, J. Johnson, J. Lees, and M. Welsh, "Deploying a wireless sensor network on an active volcano," *IEEE internet computing*, vol. 10, no. 2, pp. 18–25, 2006.

[3] M. Suzuki, S. Saruwatari, N. Kurata, and H. Morikawa, "A high-density earthquake monitoring system using wireless sensor networks," in *Proceedings of the 5th international conference on Embedded networked sensor systems*. ACM, 2007, pp. 373–374.

[4] Y. Wang, Y. Wang, X. Qi, L. Xu, J. Chen, and G. Wang, "L3sn: A level-based, large-scale, longevous sensor network system for agriculture information monitoring," *Wireless Sensor Network*, vol. 2, no. 09, p. 655, 2010.

[5] P. Dutta and D. Culler, "Practical asynchronous neighbor discovery and rendezvous for mobile sensing applications," in *Proceedings of the 6th ACM conference on Embedded network sensor systems*. ACM, 2008, pp. 71–84.

[6] A. Kandhalu, K. Lakshmanan, and R. R. Rajkumar, "U-connect: a low-latency energy-efficient asynchronous neighbor discovery protocol," in *Proceedings of the 9th ACM/IEEE international conference on information processing in sensor networks*. ACM, 2010, pp. 350–361.

[7] M. Bakht, M. Trower, and R. H. Kravets, "Searchlight: Won't you be my neighbor?" in *Proceedings of the 18th annual international conference on Mobile computing and networking*. ACM, 2012, pp. 185–196.

[8] W. Sun, Z. Yang, K. Wang, and Y. Liu, "Hello: A generic flexible protocol for neighbor discovery," in *INFOCOM, 2014 Proceedings IEEE*. IEEE, 2014, pp. 540–548.

[9] L. Chen, R. Fan, K. Bian, M. Gerla, T. Wang, and X. Li, "On heterogeneous neighbor discovery in wireless sensor networks," in *Computer Communications (INFOCOM), 2015 IEEE Conference on*. IEEE, 2015, pp. 693–701.

[10] M. J. McGlynn and S. A. Borbash, "Birthday protocols for low energy deployment and flexible neighbor discovery in ad hoc wireless networks," in *Proceedings of the 2nd ACM international symposium on Mobile ad hoc networking & computing*. ACM, 2001, pp. 137–145.

[11] S. Vasudevan, D. Towsley, D. Goeckel, and R. Khalili, "Neighbor discovery in wireless networks and the coupon collector's problem," in *Proceedings of the 15th annual international conference on Mobile computing and networking*. ACM, 2009, pp. 181–192.

[12] L. You, Z. Yuan, P. Yang, and G. Chen, "Aloha-like neighbor discovery in low-duty-cycle wireless sensor networks," in *Wireless Communications and Networking Conference (WCNC), 2011 IEEE*. IEEE, 2011, pp. 749–754.

[13] T. Song, H. Park, and S. Pack, "A probabilistic neighbor discovery algorithm in wireless ad hoc networks," in *Vehicular Technology Conference (VTC Spring), 2014 IEEE 79th*. IEEE, 2014, pp. 1–5.

[14] G. Bianchi, L. Fratta, and M. Oliveri, "Performance evaluation and enhancement of the csma/ca mac protocol for 802.11 wireless lans," in *Personal, Indoor and Mobile Radio Communications, 1996. PIMRC'96., Seventh IEEE International Symposium on*, vol. 2. IEEE, 1996, pp. 392–396.

[15] Y. Wang, W. Fu, and D. P. Agrawal, "Gaussian versus uniform distribution for intrusion detection in wireless sensor networks," *IEEE Transactions on Parallel and Distributed systems*, vol. 24, no. 2, pp. 342–355, 2013.

[16] J.-R. Jiang, Y.-C. Tseng, C.-S. Hsu, and T.-H. Lai, "Quorum-based asynchronous power-saving protocols for ieee 802.11 ad hoc networks," *Mobile Networks and Applications*, vol. 10, no. 1-2, pp. 169–181, 2005.

[17] W.-S. Luk and T.-T. Wong, "Two new quorum based algorithms for distributed mutual exclusion," in *Distributed Computing Systems, 1997., Proceedings of the 17th International Conference on*. IEEE, 1997, pp. 100–106.

[18] C. Ding, D. Pei, and A. Salomaa, *Chinese remainder theorem: applications in computing, coding, cryptography*. World Scientific, 1996.

[19] V. Daiya, J. Ebenezer, S. S. Murty, and B. Raj, "Experimental analysis of rssi for distance and position estimation," in *Recent Trends in Information Technology (ICRTIT), 2011 International Conference on*. IEEE, 2011, pp. 1093–1098.

[20] T. Moscibroda and R. Wattenhofer, "The complexity of connectivity in wireless networks." in *Infocom*, 2006, pp. 1–13.

[21] L. Wang, P.-J. Wan, and W. Washington, "Connectivity of multihop wireless networks with log-normal shadowing," *Wireless Networks*, vol. 21, no. 7, pp. 2279–2292, 2015.

[22] E. Lebhar and Z. Lotker, "Unit disk graph and physical interference model: Putting pieces together," in *Parallel & Distributed Processing, 2009. IPDPS 2009. IEEE International Symposium on*. IEEE, 2009, pp. 1–8.

[23] A. Dunkels, "The contikimac radio duty cycling protocol," 2011.

[24] L. F. Van Hoesel and P. J. Havinga, "A lightweight medium access protocol (lmac) for wireless sensor networks: Reducing preamble transmissions and transceiver state switches," 2004.

[25] A. Flammini, D. Marioli, E. Sisinni, and A. Taroni, "A real-time wireless sensor network for temperature monitoring," in *Industrial Electronics, 2007. ISIE 2007. IEEE International Symposium on*. IEEE, 2007, pp. 1916–1920.

[26] R. Motwani and P. Raghavan, "Randomized algorithms," *Acm Sigact News*, vol. 26, no. 3, pp. 48–50, 1995.

[27] K. Wang, X. Mao, and Y. Liu, "Blinddate: A neighbor discovery protocol," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 4, pp. 949–959, 2015.