Performance Limits of Neighbor Discovery in Wireless Networks

Philipp H. Kindt and Samarjit Chakraborty, Fellow, IEEE

Abstract—Neighbor Discovery (ND) is the procedure employed by wireless devices to establish a first contact. All ND protocols involve devices sending beacons, and also listening for them. Protocols differ in terms of how the beacon transmissions and reception windows are scheduled, and the device sleeps in between consecutive transmissions and reception windows in order to save energy. A successful discovery constitutes a sending device’s beacon coinciding with a receiving device’s reception window. The goal of all ND protocols is to minimize the discovery latency. In spite of the ubiquity of ND protocols and active research on this topic for over two decades, the basic question “Given a power budget, what is the minimum guaranteed ND latency?”, however, has still remained unanswered. This paper is on the best-achievable ND latency for a given power budget between a pair of devices. In order to compute this lower bound, we introduce a concept called coverage maps, that allows us to analyze the ND procedure in a protocol-independent manner. Using it, we derive discovery latencies for different scenarios, e.g., when both devices have the same or different power budgets. We also show that some existing protocols can be parametrized such that they perform optimally. Our results are restricted to the case when a few devices discover each other at a time, as is the case in most real-life scenarios, while scenarios with large numbers of devices need further study.

Index Terms—Asynchronous communication, personal area networks, wireless networks.

I. INTRODUCTION

In ENERGY-CONSTRAINED mobile devices, radios are duty-cycled and wake up only for short amounts of time for carrying out the necessary communication and then go back to a sleep mode. While such duty-cycled communication schemes are easy to realize when the clocks of all devices are synchronized and their wake-up schedules are known by all participants of the network, asynchronous communication (i.e., communication without synchronized clocks) remains a challenging problem. One of the most important asynchronous procedures is establishing a first contact between different devices, which is referred to as neighbor discovery (ND).

Neighbor Discovery: ND is used by a device for detecting other devices in range. This could be for clock synchronization and establishing a connection, after which more data can be exchanged in a synchronous fashion. Efficient ND is characterized by achieving the shortest possible discovery latency for a given energy budget. Towards this, a large number of ND protocols have been proposed till date, see [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47]. Among these [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], for example, are concerned with deterministic discovery. Here, given the protocol parameters, an upper bound on the discovery latency between a pair of devices can be determined. The problem of pairwise discovery between two devices is of fundamental importance, since in many scenarios, devices join the network gradually and only a central device and the newly joining one carry out the discovery procedure simultaneously. Moreover, the process of discovering multiple devices always relies on pairwise ND. Over the years, successive ND protocols have improved their discovery latencies for given energy budgets. For example, the Griasdi protocol proposed in 2017 claims to achieve by 87% lower worst-case latencies than Searchlight-Striped [9] that was proposed in 2012. However, despite the significant attention the ND problem has received over the past 15+ years, the fundamental question of what is the theoretically lowest possible discovery latency that any ND protocol could guarantee for a given energy budget still remained unanswered. We next provide a technical description of what is a ND protocol and how the properties of such a protocol relate to optimal performance.

Protocols for ND: The performance (e.g., worst-case discovery latency, energy consumption, etc.) of the ND procedure is fully determined by the wake-up schedules for transmission and reception of two devices discovering each other. While only very few constraints limit the set of feasible schedules for transmission and reception for technical reasons (e.g., no transmission and reception can be scheduled at the same time), existing protocols for ND reduce this design space considerably. Any protocol for ND is essentially a “construction plan” for creating a set of schedules for transmission and reception, and only schedules allowed by this plan can be realized. Moreover, all known protocols provide one or multiple parameters for adjusting the resulting schedules to practical needs, e.g., to the energy budgets of the radios executing the schedules.

Obviously, the highest performance of a particular ND protocol is achieved for certain, protocol-specific, optimal configurations, and it is often not trivial to identify these Pareto-points. However, even when a specific ND protocol is configured optimally, this does not mean that the resulting
performance cannot be superseded by a different ND protocol. In fact, the construction plan the ND protocol implies might not result in an optimal set of wake-up schedules, leading to a non-optimal performance even when the parameters that lead to the highest performance have been chosen. Similarly, a protocol that actually results in an optimal set of wake-up schedules for some parametrizations does not necessarily perform optimally, when a different, non-optimal parametrization is used. We next discuss the difficulties in assessing the performance of ND protocols.

**Performance of ND Protocols:** In the absence of a protocol-agnostic bound on the discovery latency, the performance evaluations of different ND protocols have often been very subjective. The results of such evaluations relied on the choice of protocols, their parametrizations and the assumed setups. Hence, while a certain protocol might outperform others in such a comparison, it might perform differently if the parametrization or setup is changed. In addition, most known protocols, e.g., [7], [9], [14], subdivide time into multiple slots and are hence referred to as slotted. A device sleeps in most slots, whereas some slots are active and used for communication. Discovery occurs once two active slots overlap in time. Here, performance is quantified in terms of the worst-case number of slots until discovery is guaranteed. Though a certain protocol could perform better than another in terms of the number of slots, such comparisons are heavily dependent on the supported range of slot lengths. As a result, such comparisons in terms of slots, and not directly in terms of time, are often not meaningful. Moreover, despite slotted protocols having been studied thoroughly in the literature, many protocols that are frequently used in practice, e.g., Bluetooth Low Energy (BLE), do not rely on a slotted paradigm. They schedule reception windows and beacon transmissions with periodic intervals and offer three degrees of freedom that can be configured [48] (viz., the periods for reception and transmission, and the length of the reception window). It has been shown that some parametrizations for ND in BLE networks proposed by official specifications [49] might lead to a non-optimal performance [18]. This has raised the interest to fully understand such slotted ND procedures. In particular, finding beneficial parametrizations for periodic interval-based protocols has been studied in the literature recently, e.g., in [16], [17], and [18]. However, prior to this work, it was neither clear whether the proposed parametrizations are actually optimal, nor how their performance compares to slotted protocols. In summary, despite the large volume of available literature, it has not been possible to meaningfully assess and classify the performance of ND protocols in a purely objective fashion.

**Overview of this paper:** In this paper, we study the fundamental limits of pairwise, deterministic ND. In particular, we establish a relationship between the optimal discovery latency, channel utilization (and hence beacon collision rate) and power budget. No pairwise ND protocol can achieve lower discovery latencies than the ones established in this paper. The resulting bounds not only give important insights into the design of ND protocols, but will serve as a baseline for more objective performance comparisons. Surprisingly, our analysis shows that some recently proposed protocols actually perform optimally and cover parts of the latency/channel utilization/power budget Pareto front. The optimality results of such protocols were not known previously. We show in this paper how to modify such protocols to cover the entire Pareto-front. The coverage of the entire Pareto front implies that there is no further potential for improvement. However, there is still potential to improve the robustness against beacon collisions, which might occur frequently when many devices carry out ND simultaneously. The result that practical protocols can actually reach the performance given by the entire Pareto front also shows that our proposed bounds are tight.

**Basic principles of ND:** In general, a radio can either be in a sleep state, listen to the channel or transmit a beacon. Hence, the basic building blocks of a ND protocol are given by these three operations, and any ND protocol can be represented as a sequence of them. For higher power-budget, the number of beacons and/or the number or lengths of reception windows can be increased. A discovery procedure is successful, once a beacon coincides with a reception window on another device. Since the design space of all possible reception and transmission patterns allows for an infinite number of possible configurations, determining the optimal pattern and its performance through any form of exhaustive search or numerical method is not possible. Furthermore, as outlined above, most work on ND has focused on slotted protocols and therefore studied only a small part of the design space. As a result, the problem of assessing the optimal performance of the ND procedure has remained unsolved prior to this work.

**ND Scenarios:** For different scenarios, the ND problem appears in different forms, and we provide bounds on the discovery latency for many of them. In all of them, we restrict ourselves to procedures in which one pair of devices attempt to discover each other. First, it is obvious that if two devices $E$ and $F$ both have the same beacon and reception patterns, their discovery properties are symmetric. This implies that device $E$ discovers device $F$ with the same worst-case latency for a given power budget as $F$ discovering $E$. Several publications, e.g., [7], [18], [21], have studied this special case of symmetric power budgets, for which we present a bound on the discovery latency. If both devices run different patterns (for example, due to different budgets), the discovery properties are asymmetric. For the asymmetric case, we provide a bound on the discovery latency when each device is aware of the other device’s configuration. The problem of two devices being allowed to modify their patterns autonomously during operation is also relevant. Whether the bounds we present for the asymmetric case can also be achieved when one device is not aware of the patterns of its opposite one is unclear and needs further study.

Another important question we answer in this paper is the partitioning of the power budget of a device. The power budget of a device is characterized by the fraction of time it is active. On the other hand, channel utilization is the fraction of time a device occupies the channel, which is between zero and its duty-cycle, i.e., the total amount of time a device is awake. Beacon collision rates are solely determined by the channel utilizations of the devices in range. For the case when the channel-utilization (and hence collision rate) is unconstrained, we derive the ratio between transmission and reception times that minimizes the discovery latency. In the case of many devices discovering each other, the channel utilization of each
In this paper, we therefore not only derive bounds on the discovery latency that any protocol can guarantee for a given power budget, but also for the case where both the power budget and the maximum channel-utilization are provided. In summary, this work makes the following contributions.

Technical Contributions: We present the following bounds on the discovery latency of deterministic ND protocols.

1) The lowest discovery latency any symmetric and asymmetric pairwise ND protocol can guarantee for a given power budget.

2) A discovery latency bound for the case where the channel utilization is additionally constrained.

3) Bounds for the following cases in which two devices \( E \) and \( F \) discover each other. (a) Only \( E \) discovers \( F \), whereas \( F \) cannot discover \( E \). (b) Either \( E \) discovers \( F \) or \( F \) discovers \( E \), but both discovering each other is not possible. (c) Both \( E \) and \( F \) mutually discover each other.

We further study the relation between our bounds and previously known ones [10], [11], [20], [21], which are all limited to slotted protocols. These bounds are given in terms of a worst-case number of slots until discovery is guaranteed, where the discovery latency also depends on the slot length. However, how small a slot length can be is difficult to answer, while it is known that slot lengths cannot be made arbitrarily small. Therefore, the lowest possible discovery latencies of slotted protocols in terms of time have not been derived, which we address in this paper. Finally, while most previous work has focused on slotted protocols, we show that when the channel utilization is unconstrained, only slotted protocols can perform optimally, whereas slotted ones cannot. This result is important because in many IoT scenarios, devices join the network gradually, and only a pair of devices participate in the ND procedure at the same time. Here, channel utilization is therefore often not of concern.

Importance of Performance Bounds for ND: In addition to their theoretical importance, and the new insights provided by our results, they have several practical applications. Our results help in understanding how to configure existing protocols such as BLE to obtain low latencies and energy consumption. This will become increasingly important for battery-powered IoT devices, and a large number of existing BLE devices can benefit from increased battery lifetimes. Second, our results help in developing practical protocols that are tailored to a certain application, while providing latencies beyond what is possible using already deployed protocols, e.g., BLE. For example, contact tracing using smartphones or custom wearables has received significant attention in the course of the COVID-19 pandemic in 2020 [50]. Here, devices for contact tracing carry out ND continuously as their main mode of operation, and therefore, efficient ND protocols are of crucial importance.

Organization of the paper: The rest of this paper is organized as follows. In Section II, we present related work on discovery latency bounds of ND protocols. Next, in Section III, we provide a formal description of a generic ND procedure. Based on this, in Section IV, we derive a list of properties that deterministic ND protocols need to fulfill. We derive generic latency bounds for such deterministic protocols in Section V. Finally, in Section VI, we relate the bounds of multiple existing ND protocols to the bounds obtained in this paper. We also show how to extend existing protocols, such that every point on the Pareto-front spanned by the worst-case latency, power budget and channel utilization can be reached. Throughout this paper, we make a couple of simplifying assumptions. These assumptions are only for the ease of exposition only and are relaxed in the appendix. The appendix, which is available as supplemental material, also studies non-ideal radios, presents additional scenarios, provides detailed derivations, and contains a comprehensive table of symbols.
Generic Approaches: Unlike the works described above that were specific to slotted or PI-based protocols, protocol-agnostic bounds were first presented in [52] and [53]. In particular, they give an asymptotic latency bound in the form of $O(d)$, where $d$ is “the discretized uncertainty period of the clock shift between the two processors” [52]. Hence, this bound depends on the degree of asynchrony between the clocks of a sender and a receiver. First, the asymptotic nature of such a bound is very different from the concrete power budget, which are of direct practical relevance. For these reasons, the bounds from [52] and [53] are not comparable to those that have been more commonly pursued, and also to those presented in this paper.

This Paper: This paper is an extension of [54], where generic, protocol-agnostic bounds on the lowest discovery latency that can be achieved for a given power budget were presented. Compared to [54], this paper contains multiple extensions. First, in [54], a bound for asymmetric ND that is valid for only a restricted set of power budgets is extended to all power budgets in this paper. Similarly, we in the appendix, which is available as supplemental material, relax almost all assumptions that were present in [54]. We furthermore study the impact of clock skew in the appendix. In addition, we describe a protocol with optimal performance that can obey a certain constraint on the channel utilization. We furthermore also study the mean latencies achieved by protocols with optimal worst-case latencies. Similarly, we study the mean latencies of probabilistic protocols, which cannot guarantee any bounded latencies. In addition, compared to [54], we have revised and extended multiple sections to make the theory easier to understand. In summary, this paper covers the topic of performance bounds in a more comprehensive and accessible fashion compared to [54]. While this paper establishes the theoretical foundations of optimal ND, an experimental evaluation of such an optimal protocol is presented in [18].

III. NEIGHBOR DISCOVERY PROTOCOLS

A. Definition

In this section, we formally define the ND procedure and its associated properties.

Definition 1 (Reception Window Sequence): Let the time windows during which a device listens to the channel be given by the tuples $c_1 = (t_1, d_1)$, $c_2 = (t_2, d_2)$, $c_3 = (t_3, d_3)$, ..., where each reception window $c_i$ starts at time $t_i$ and ends $d_i$ time-units later (see Figure 1). A reception window sequence $C = c_1, c_2, ..., c_n$ could be of finite or infinite length. In this paper, for simplicity of notation, we refer to such finite length sequences by $C$ and infinite length sequences by $C_{\infty}$.

For the simplicity of exposition, throughout this paper, we always assume that any $C_{\infty}$ is an infinite concatenation of some finite length sequence $C$. For such $C_{\infty}$, we define $n_C = |C|$ (i.e., the number of windows contained in $C$). Further, we denote the time between the ends of two consecutive instances of $C$ as the reception period $T_C$. All our bounds remain valid also for sequences $C_{\infty}$ that are not given by concatenating the same $C$, as we show in Appendix C.

We assign a time-axis to every instance of $C$. For convenience, which will become clear later, the origin of time in a certain instance of $C$ will start at the end of the last reception window of the previous instance, as depicted in Figure 1. In this figure, $C$ consists of three reception windows (i.e., $c_1, c_2, c_3$), and three concatenated instances of $C$ are shown. For example, the origin of the time-axis for Instance 2 lies at the end of $c_3$ in Instance 1.

Definition 2 (Beacon Sequence): A sequence of beacons $B = b_1, b_2, ..., b_m$ sent at the time-instances $\tau_1, \tau_2, ..., \tau_m$, as depicted in Figure 2, is called a beacon sequence of length $m$. The transmission durations of these beacons are given by $\omega_1, \omega_2, ..., \omega_m$. A sequence of infinite length (i.e., $m \rightarrow \infty$) is denoted by $B_{\infty}$.

We denote infinite length beacon sequences $B_{\infty}$ that are given by concatenations of a finite beacon sequence $B$ as repetitive beacon sequences. In such repetitive sequences, $m_B = |B|$ and the time between the endings of two consecutive instances of $B$ is given by $T_B$. We will prove that all beacon sequences that optimize the relevant metrics of a ND procedure are repetitive when the corresponding reception window sequence is also repetitive.

We indicate an arbitrary shorter sequence $B'$ to be a part of a longer sequence $B$ by using the notation $B' \subseteq B$. For example, in Figure 2, $B' = b_2, b_3, b_4, b_5, b_6 \subseteq B$. Further, the time between the beginnings of beacon $b_i$ and beacon $b_{i+1}$ is called the beacon gap $\lambda_i$. It is $\lambda_i = \tau_{i+1} - \tau_i$.

Definition 3 (ND Protocol): A tuple of an infinite beacon and reception window sequence $(B_{\infty}, C_{\infty})$ is called a ND protocol.

In this paper, unless explicitly stated, we assume that $B_{\infty}$ and $C_{\infty}$ stem from two different devices $E$ and $F$. When it is necessary to explicitly specify the device that a sequence is scheduled on, we use the notation $B_{E,\infty}$ or $C_{E,\infty}$, where $E$ and $F$ refer to device $E$ or $F$ respectively. We also apply this notation to reception windows and beacons, e.g., $\omega_{E,1}$ refers to beacon 1 on device $E$ and $c_{F,1}$ refers to reception window 1 on device $F$. Depending on the considered scenario, every device can either run a beacon or reception window sequence only, which would lead to unidirectional discovery, or simultaneously run a pair of sequences to support bidirectional discovery. The bidirectional case can be regarded as two independent unidirectional procedures running simultaneously. Hence, we can in the following think of the ND procedure
as one device discovering one other device unidirectionally, without loss of generality. The most important properties of a ND protocol are its worst-case latency $L$, its power budget $\eta$, and its channel utilization $\beta$, as defined next.

**Definition 4 (Worst-Case Latency):** Given two devices $E$ and $F$, where $E$ runs an infinite beacon sequence and $F$ an infinite reception window sequence, the worst-case latency $L$ is the earliest possible time after which an overlap of a beacon from $E$ with a reception window of $F$ is guaranteed, measured from the point in time both devices come into the reception range.

The worst-case latency is widely accepted as the performance metric of deterministic ND protocols.

**Definition 5 (Duty-Cycle and Power Budget):** The transmission duty-cycle $\beta$ of a device is the fraction of time it spends for transmission, whereas the reception duty-cycle $\gamma$ is the fraction of time spent for reception. Therefore, the total power budget $\eta$ is given as a weighted sum $\eta = \gamma + \alpha \beta$, where $\alpha$ is the ratio of transmission and reception powers, i.e., $\alpha = \frac{\nu_T}{\nu_R}$. For a radio running a tuple of sequences $(B_\infty, C_\infty)$, it is:

$$\beta = \lim_{m \to \infty} \sum_{i=1}^{m-1} \frac{\omega_i}{T_m - t_1}, \quad \gamma = \lim_{n \to \infty} \sum_{i=1}^{n-1} \frac{d_i}{t_n - t_1}, \quad \eta = \alpha \beta + \gamma.$$

(1)

The transmission duty-cycle $\beta$ is the same as the channel utilization. The power budget $\eta$ directly corresponds to the power consumption of an ideal radio. Non-ideal radios incur additional overheads, e.g., turnaround times to switch between transmission and reception. Such non-idealities are discussed in the appendix. When the current consumption for reception is equal to that for transmission (i.e., $\alpha = 1$), the power budget $\eta$ is equal to the sum of reception and transmission duty-cycles. Furthermore, note that the power budget of a receiver is equal to $\gamma$, and the power budget of a transmitter is characterized by $\beta$, as $\alpha$ is immutable for a considered radio. The duty-cycles and power budget of a tuple of sequences $B_\infty, C_\infty$ that are concatenations of finite length sequences $B$ and $C$, respectively, can be computed as follows.

$$\beta = \frac{\sum_{i=1}^{m_B} \omega_i}{T_B}, \quad \gamma = \frac{\sum_{i=1}^{m_C} d_i}{T_C}, \quad \eta = \alpha \beta + \gamma.$$

(2)

For example, a protocol that transmits one beacon with a period of $T_B$ and listens to the channel for $10 \cdot \omega$ time-units with a period of $T_C$ has a transmission duty-cycle of $\beta = \omega / T_B$, and a reception duty-cycle $\gamma = (10 \cdot \omega) / T_C$.

**B. Beacon Length**

A beacon needs to be transmitted entirely within a reception window of a receiving device for being received successfully. Each beacon has a certain transmission duration $\omega_i$, and if the beacon transmission starts after the last $\omega_i$ time-units of a reception window (cf. after the start of the hatched area in Figure 3a)), it cannot be received successfully. Nevertheless, for the sake of simplicity of exposition, we assume that any overlap between a beacon and a reception window leads to a successful discovery. We further assume that all beacons have the same length $\omega$ and neglect the contribution of the transmission duration of the first successfully received beacon to the worst-case latency. We study the relaxation of these assumptions in the appendix.

**IV. DETERMINISTIC BEACON SEQUENCES**

A device $F$ can successfully discover another device $E$ only if $E$ sends a beacon during one of the reception windows of $F$. We refer to the other direction as $E$ discovering $F$. In what follows, we first consider $F$ discovering $E$ only, and later generalize it towards mutual discovery. On device $E$, let $B' = b_1, b_2, \ldots$ be a subsequence of $B_\infty$. Here, we will always assume that $b_1$ is the first beacon that is in range of a remote device $E$. This is because any prior beacons of $B_\infty$, when $E$ is not within the range of $F$, are not relevant for ND. Further, let $F$ run an infinite reception window sequence $C_\infty$. Though $B_\infty$ and hence $B' \subseteq B_\infty$ could be of infinite length, let us think of $B'$ as a fixed-length sequence. This assumption is valid because in the case of a successful discovery, beacons that are sent thereafter are no longer relevant for the discovery procedure. Now recall that the reception windows of $C_\infty$ are formed by concatenations of a finite sequence $C$ and every instance of $C$ has its own time origin, as defined by Definition 1 (cf. Figure 1). The first beacon $b_1$ in $B'$ lies within a certain instance of $C$ and has a certain (random) offset $\Phi$ from the time origin of this instance of $C$. This is depicted in Figure 3b), which shows an infinite reception window sequence consisting of concatenations of $C = c_1, c_2, c_3$, of which one full instance is depicted. In addition, the figure contains the last reception window $c_3$ of the preceding instance and the first reception window $c_1$ of the succeeding one. Further, three beacons $b_0, b_1$ and $b_2$ are shown, of which only $b_1$ and $b_2$ are sent in range. Here, $B'$ consists of $b_1, b_2$ and some later beacons that are not shown in the figure. Beacon $b_1$ falls into the depicted instance of $C$ and has an offset of $\Phi_1$ time-units from its origin.

For some valuations of $\Phi_1$, at least one beacon of $B'$ will coincide with a reception window of $C_\infty$. For other valuations of $\Phi_1$, there might be no beacon in $B'$ that coincides with any reception window of $C_\infty$, irrespective of the length of $B'$. In other words, whether a beacon coincides with a reception window or not solely depends on which amount of time $B'$ and $C_\infty$ are “shifted” against each other, and this amount of “shift” is given by the random value $\Phi_1$. If a coinciding pair of a beacon and a reception window exists for all possible offsets $\Phi_1$, the tuple $(B', C_\infty)$ guarantees discovery within a bounded amount of time and hence realizes deterministic ND.
We, in the following, formalize the properties that such a tuple \((B', C_\infty)\) needs to fulfill for guaranteeing discovery.

A. Coverage and Determinism

A tuple \((B', C_\infty)\), along with \(\Phi_1\), is depicted in Figure 3b). For a given \((B', C_\infty)\), it is obvious that the offset \(\Phi_1\), which is a measure of the shift between \(B'\) and \(C_\infty\), solely determines whether a beacon in \(B'\) coincides with a reception window in \(C_\infty\) or not. The time after which such a coincidence takes place, and hence the discovery latency, is also determined by \(\Phi_1\). For which values of \(\Phi_1\) will beacon \(b_1\) fall into one of the reception windows? Clearly, these are given by the set 

\[
\Omega_1 = \{(t_1, t_1 + d_1), [t_2, t_2 + d_2], \ldots\} \quad \text{(cf. Figure 3b)}.
\]

In other words, \(\Phi_1\) lies within any interval belonging to \(\Omega_1\), then \(b_1\) is successfully received. Similarly, if \(\Phi_2\) is the offset of \(b_2\), then for \(\Phi_2\) belonging to any interval in \(\Omega_1\), \(b_2\) will be successfully received (see Figure 3b). Now, what are the offsets \(\Phi_1\) of \(b_1\), such that beacon \(b_2\) is successfully received? These are given by the set 

\[
\Omega_2 = \{\{t_1 - \lambda_1, t_1 + d_1 - \lambda_1\}, [t_2 - \lambda_1, t_2 + d_2 - \lambda_1], \ldots\},
\]

where \(\lambda_1\) is the time-distance between the beacons \(b_1\) and \(b_2\), as already defined in Section III (see Figure 3b). Therefore, \(\Omega_2\) is obtained by shifting all elements of \(\Omega_1\) by \(\lambda_1\) time-units to the left. Similarly, \(\Omega_3\) is 

\[
\Omega_3 = \{\{t_1 - (\lambda_1 + \lambda_2), t_1 + d_1 - (\lambda_1 + \lambda_2), [t_2 - (\lambda_1 + \lambda_2), t_2 + d_2 - (\lambda_1 + \lambda_2)], \ldots\}.
\]

Then \(\Omega_k\) for 

\[
k = 3, 4, 5, \ldots \text{ is similarly defined as}
\]

\[
\Omega_k = \left\{\begin{array}{l}
\{t_1 - \sum_{i=1}^{k-1} \lambda_i, t_1 + d_1 - \sum_{i=1}^{k-1} \lambda_i\}, \\
\{t_2 - \sum_{i=1}^{k-1} \lambda_i, t_2 + d_2 - \sum_{i=1}^{k-1} \lambda_i\}, \ldots\}
\right\}.
\]

Whenever any offset becomes negative, an appropriate multiple of \(T_C\) needs to be added to it, such that it lies between 0 and \(2T_C\) and hence falls into the considered instance of \(C\) \(\subseteq C_\infty\). Now consider a beacon sequence \(B' = \{b_1, \ldots, b_m\}\) of length \(m\). If \(\Phi_1\) belongs to any interval in \(\Omega_1 \cup \Omega_2 \cup \ldots \cup \Omega_m\), then one beacon from \(B'\) will be successfully received. We now extend this result and define a coverage map, which can be used to reason about valuations of the initial beacon offset \(\Phi_1\) that lead to successful discovery.

1) Coverage Maps: A coverage map is a formal representation of all offsets \(\Phi_1\) for which any beacon in \(B'\) overlaps with a reception window in \(C_\infty\). It also allows for a graphical representation, from which several properties of the tuple \((B', C_\infty)\) can be easily understood.

Recall that \(C_\infty\) is a repeated concatenation of a sequence of reception windows \(C\) (i.e., \(C_\infty = C C C \ldots\)). Now, we need to be able to specify specific instances of \(C\) within \(C_\infty\). For this purpose, let us consider a simple example where \(C\) has two reception windows \(X\) and \(Y\), and \(C_\infty\) is therefore given by \(C_\infty = XXYYXXY \ldots\). In order to distinguish between different instances of these reception windows, we will denote 

\[
C_\infty = X_0 Y_0 X_1 Y_1 X_2 Y_2 \ldots
\]

The reception windows \(X_i\) and \(X_{i+1}\), as well as \(Y_i\) and \(Y_{i+1}\), are \(T_C\) time-units apart (see Figure 4a) and also Figure 1).

Figure 4a shows a sequence of beacons \(B' = b_1, \ldots, b_7\) from a transmitting device. Below, two reception windows \(X_0, Y_0\) from a receiving device are depicted, together with their periodic repetitions \(X_1, Y_1\), which are \(T_C\) time-units later. Again, \(b_1 \in B'\) has a certain random offset \(\Phi_1\) from the origin of \(C\). Figure 4b) shows the coverage map for the sequences in Figure 4a).

Definition 6 (Covered): An offset \(\Phi_1\) is covered, if at least one beacon in \(B'\) coincides with any reception window in \(C_\infty\) for this offset.

Given the parameters of \((B', C_\infty)\), the construction of a coverage map as in Figure 4b), is straightforward. We believe that the notion of such a coverage map and its use go beyond deriving latency bounds as done in this paper. It would also be useful for analyzing and optimizing various kinds of different ND protocols, including already known ones. From coverage maps, we can derive the following properties.

- **Beacon-to-beacon discovery latency** \(\lambda^*\): For a given offset \(\Phi_1\), let \(\lambda^*(\Phi_1)\) be the latency measured from the transmission time of the first beacon in range, to the first time a beacon is successfully received. In Figure 4, \(\lambda^*(\Phi_1) = \tau_1 - \tau_2 = \lambda_k = \sum_{i=1}^{k-1} \lambda_i\), where \(\lambda_1\) is the smallest row number in which \(\Phi_1\) is covered. For example, for an offset \(\Phi_1\) slightly above 0 (i.e., an offset within the highlighted frame in Figure 4b)), the beacon-to-beacon discovery latency will be \(\lambda^* = \tau_3 - \tau_1\), since \(b_3\) is the earliest successful beacon for this offset.

- **Determinism**: By ensuring that all possible initial offsets are covered by at least one beacon, we can guarantee that \(B'\) is deterministic with respect to \(C_\infty\) (see next section for a formal definition of determinism).

- **Redundancy**: For certain valuations of \(\Phi_1\), one can see in Figure 4b) that a beacon will be received by multiple reception windows. For example, for values of \(\Phi_1\) within the shaded frame, beacons \(b_3\) and \(b_7\) will be received by the windows \(X_1\) and \(X_2\), respectively. By integrating over the length of all reception windows for which such duplicate receptions happen, we can quantify the degree of redundancy of a tuple \((B', C_\infty)\).

2) Determinism: Recall that protocols that can guarantee discovery for every possible initial offset are called deterministic. This is formalized below. In particular, we distinguish between a beacon sequence \(B'\) and a protocol \((B_\infty, C_\infty)\) that can result in such a sequence.

Definition 7 (Deterministic ND Protocol): A beacon sequence \(B'\) is deterministic in conjunction with an infinite
reception window sequence \( C_\infty \), if all possible initial offsets \( \Phi_1 \) are covered by the tuple \((B', C_\infty)\). A ND protocol \((B_\infty, C_\infty)\) is deterministic, if for all \( i \), \( B'_i = b_i, b_{i+1}, b_{i+2}, \ldots \) is a deterministic beacon sequence. 

Hence, deterministic ND protocols \((B_\infty, C_\infty)\) always guarantee a bounded discovery latency, no matter when a beacon of \( B_\infty \) comes within the range of a receiving device.

**Lemma 1:** If a beacon sequence \( B' \) covers all offsets \( \Phi_1 \) within \([0, T_C]\), then all possible valuations of \( \Phi_1 \) are covered.

**Proof:** Let us assume that a certain range of offsets \([X, Y]\), where \( X, Y \leq T_C \), is covered by a beacon \( b_i \) in conjunction with a certain reception window \( c_j \). Since the pattern of reception windows repeats every \( T_C \) time-units, any \( \Phi_1 \in [X + T_C, Y + T_C] \) will result in \( b_i \) being received by the reception window \( c_{j+n_{BC}} \), which is \( T_C \) time-units after \( c_j \). \( \square \)

**Definition 8 (Redundant Sequences):** If any offset \( \Phi_1 \) within \([0, T_C]\) is covered by more than one beacon, then the tuple \((B', C_\infty)\) is redundant. Otherwise, \((B', C_\infty)\) is disjoint, since no intervals in the corresponding coverage map overlap.

For example, in Figure 4b), all offsets \( \Phi_1 \) are covered and hence, the corresponding tuple \((B', C_\infty)\) is deterministic. Further, since some offsets, e.g., the ones slightly above offset 0 (marked by the highlighted frame in Figure 4b)) are covered twice, it is also redundant. For the purpose of ND involving two devices, it is never beneficial to realize redundant coverage. In addition, packets for ND should contain as few data as possible, e.g., only a device ID and synchronization information. If more data is to be exchanged, doing so in a synchronized manner after an initial beacon reception is typically by orders of magnitude more energy-efficient than exploiting redundancy.

3) **Coverage:** For a tuple \((B', C_\infty)\), certain values of \( \Phi_1 \) might be covered by multiple beacons, other values by exactly one beacon and yet others by no beacons. The notion of coverage quantifies how different values of \( \Phi_1 \in [0, T_C] \) are covered. To understand this, recall that \( \Omega_i \) is a set of intervals. Let us now consider those (full or partial) intervals of \( \Omega_i \) that lie within \([0, T_C]\). The sum of the lengths of all such intervals for all \( \Omega_i \) captures a notion of coverage that we formalize below.

**Definition 9 (Coverage):** Given a tuple \((B', C_\infty)\), let a certain offset \( \Phi_1 \in [0, T_C] \) be covered by \( k \) beacons, where \( k \in \{0, 1, 2, \ldots \} \). Let us define an auxiliary function \( \Lambda^*(\Phi_1) = k \). Then, the coverage \( \Lambda \) is defined as

\[
\Lambda = \int_0^{T_C} \Lambda^*(\Phi_1) \, d\Phi_1. \tag{4}
\]

For example, in Figure 4b), if the lengths of \( X_1 \) and \( Y_1 \) are equal to unity and therefore \( T_C = 8 \), then \( \Lambda = 14 \) (since exactly 14 shifted instances of \( X_1 \) and \( Y_1 \) fall into every instance of \( T_C \), as can be seen in Figure 4b)). If \( \Lambda < T_C \), a tuple \((B', C_\infty)\) cannot be deterministic, which implies that for certain values of \( \Phi_1 \), no bounded discovery latency can be guaranteed. If \( \Lambda = T_C \), then \((B', C_\infty)\) can either be deterministic and disjoint, or else, it will be redundant and not deterministic. If \( \Lambda > T_C \), then \((B', C_\infty)\) cannot be disjoint, and may or may be not deterministic.

### B. Minimum Coverage

While \( \Lambda \) quantifies the coverage due to all beacons in \( B' \), we now quantify the coverage induced by individual beacons. 

**Theorem 1 (Coverage per Beacon):** Given a tuple \((B', C_\infty)\), every beacon \( b_i \in B' \) induces a coverage of exactly \( \Lambda = \sum_{k=1}^{n_{BC}} d_k \) time-units.

**Proof:** The first beacon \( b_i \) in \( B' \) will cover exactly those time-units for which \( b_i \) directly coincides with a reception window. The sum of such coinciding offsets is therefore \( \sum_{k=1}^{n_{BC}} d_k \) time-units. Every later beacon \( b_i \) will cover the same offsets shifted by the sum of beacon gaps \( \sum_{k=1}^{n_{BC}} \lambda_k \) to the left, which does not impact the amount of offsets covered. Since \( C_\infty \) is an infinite concatenation of a finite sequence \( C \), for every covered offset that is shifted out of the considered range \([0, T_C] \), the same amount from a later period is shifted into that range, such that each beacon \( b_i \) covers \( \sum_{k=1}^{n_{BC}} d_k \) time-units within \([0, T_C] \). \( \square \)

Using this, we can derive a minimum length of \( B' \).

**Theorem 2 (Beaconing Theorem):** Given a tuple \((B', C_\infty)\), the minimum number of beacons \( M \) a beacon sequence \( B' \) needs to consist of to guarantee deterministic discovery is:

\[
M = \left\lceil \frac{T_C}{\sum_{k=1}^{n_{BC}} d_k} \right\rceil. \tag{5}
\]

**Proof:** From Theorem 1 follows that every beacon induces a coverage of \( \Lambda = \sum_{k=1}^{n_{BC}} d_k \). For deterministic discovery, the coverage \( \Lambda \) has to be at least \( T_C \). Therefore, the number of beacons needed for deterministic ND must be at least \( \lceil T_C / \Lambda \rceil \). \( \square \)

It is worth mentioning that Theorem 2 is a necessary, but not sufficient condition for deterministic ND. The positioning of the beacons, along with their quantity, also determine whether or not a tuple \((B', C_\infty)\) is deterministic.

### V. FUNDAMENTAL BOUNDS

In this section, we derive the lower bounds on the worst-case latency that an ND protocol could guarantee in different scenarios (e.g., symmetric or asymmetric discovery). In other words, given constraints like the power budget, such a bound defines the best worst-case latency that any protocol could possibly realize. First, we consider the most simple case in which one device \( F \) runs an infinite reception window sequence \( C_{F, \infty} \) without transmitting, whereas another device \( E \) only runs an infinite beacon sequence \( B_{E, \infty} \) without ever listening. We refer to this as unidirectional beaconing.

#### A. Bound on Unidirectional Beacons

1) **The Coverage Bound:** Consider a tuple \((B', C_\infty)\), where \( B' \) consists of \( M \) beacons, and \( M \) is given by Theorem 2. Recall Theorem 2 and the subsequent discussion. If \( B' \) is disjoint and deterministic, then for every value of \( \Phi_1 \), there is exactly one beacon in \( B' \) that coincides with a reception window in \( C_\infty \) What are the beacon gaps \( \lambda \) to space such \( M \) beacons for minimizing the discovery latency?
Fig. 5. Partial sequences of an infinite beacon sequence.

The worst-case beacon-to-beacon discovery latency $l^*$, measured from the first beacon in range to the earliest successfully received one, is given by the sum of the $M - 1$ beacon gaps between these beacons. The first beacon in $B'$ is the first beacon that was sent after the transmitter came into the reception range of the receiver. To measure the worst-case discovery latency $L$, time begins when the two devices come in range, which might be earlier than the time the first beacon in $B'$ was sent. How much earlier? At most by the beacon gap $b$ that precedes $B$ in $B$. Recall that $B'$ belongs to an infinite sequence $B_\infty$. Hence, the largest worst-case latency is achieved if the sum of these $M$ beacon gaps is minimized. At the same time, all offsets in $[0, T_C]$ need to be covered exactly once for ensuring determinism.

However, the following arguments rule out such $M$ consecutive beacon gaps to be arbitrarily short. $B_\infty$ has a transmission duty-cycle $\beta$, defined by the power budget of the transmitter. Obviously, $\beta$ determines the average beacon gap $\lambda$. If the sum of certain $M$ consecutive beacon gaps becomes smaller than $M \cdot \lambda$, then the sum of a different $M$ consecutive beacon gaps within $B_\infty$ needs to exceed $M \cdot \lambda$ in order to respect the average beacon gap of $\lambda$ defined by $\beta$. Since any beacon in $B_\infty$ could be the first beacon in range, the $M$ beacons with the largest sum of beacon gaps determine the worst-case latency $L$. Hence, in an optimal $B_\infty$, every sum of $M$ consecutive beacon gaps must be equal to $M \cdot \lambda$. It is worth noting that this requirement does not necessarily require equal beacon gaps, because the above property has to hold for a specific value of $M$ given by Theorem 2. This is formalized in Lemma 2.

To illustrate the above, consider the following example. Figure 5 shows a sequence $B' = b_1, \ldots, b_7$. Here, the minimum number of $M$ of beacons for deterministic ND be equal to 4 and let the partial sequences $(b_1, \ldots, b_4), (b_2, \ldots, b_5), (b_4, \ldots, b_7)$ be deterministic. Consider the sequence $b_1, \ldots, b_4$. Let us assume that $b_4$ would be sent somewhat earlier than depicted. Then, by decreasing $\lambda_4$, the beacon gap $\lambda_4$ would increase accordingly, and though the sequence $b_1, \ldots, b_4$ would result in a shorter discovery latency for all possible offsets, the sequence $b_4, \ldots, b_7$ would lead to a larger worst-case latency. The above observations are formalized below.

**Theorem 3 (Coverage Bound):** The lowest worst-case latency that can be guaranteed by a tuple $(B_\infty, C_\infty)$ is:

$$L = \left\lceil \frac{T_C}{\sum_{i=1}^{n_C} d_i} \right\rceil \omega \beta.$$

**Proof:** Consider a sequence $B' = b_1, \ldots, b_m$ with $m \gg M$. In $B'$, if any sum of $M$ consecutive beacon gaps is lower than $M \cdot \lambda$, then the sum of a different $M$ consecutive beacon gaps will exceed $M \cdot \lambda$ and will define $L$. Since this is true for every $m$, it also holds for $B_\infty$. The mean beacon gap is given by $\lambda = (\tau_m - \tau_s)/(m-1)$ and the worst-case latency by $L = M \cdot \lambda$. Expressing the mean beacon gap by the duty-cycle for transmission (cf. Equation 1) and expanding $M$ using Theorem 2 leads to Equation 6. We write an equality sign “$\equiv$” for $L$ (instead of e.g., $\leq$), since $L$ always refers to the bound itself and not to the outcome of a single discovery procedure. We will show in Section VI that our proposed bounds are tight and can hence actually be reached by practical protocols.

**Lemma 2 (Repetitive Beacon Sequences):** Given a repetitive $C_\infty$, every $B_\infty$ that guarantees the lowest possible worst-case latency is repetitive, with a period of $m_B = M$ beacons or $T_B = M \cdot \frac{\tau}{\gamma}$ time-units.

2) **Optimal Reception Window Sequences:** We know that in an optimal beacon sequence, the sum of every $M$ consecutive beacon gaps is equal to $T_B$. The corresponding reception window sequence must be designed such that all offsets in $[0, T_C]$ are covered. While there can be multiple such $C_\infty$ for a given $B_\infty$, all optimal ones fulfill the following property.

**Theorem 4 (Overlap Theorem):** Consider a tuple $(B_\infty, C_\infty)$, which guarantees a certain worst-case latency $L$. Every $C_\infty$ that achieves this latency with the lowest possible reception duty-cycle $\gamma$ fulfills the following property.

$$T_C = k \cdot \sum_{i=1}^{n_C} d_i, \quad k \in \mathbb{N}. \quad (7)$$

**Proof:** We assume that $T_C$ is equal to $k \cdot \sum_{i=1}^{n_C} d_i - \Delta$, where $k$ is an integer and $\Delta \in [0, \sum_{i=1}^{n_C} d_i]$. Theorem 3 implies the same worst-case latency for all values of $\Delta$, since the ceiling function in Equation 6 does not change $L$. With $T_C = k \cdot \sum_{i=1}^{n_C} d_i - \Delta$, $\gamma$ is given by (cf. Equation 2):

$$\gamma = \frac{\sum_{i=1}^{n_C} d_i}{k \cdot \sum_{i=1}^{n_C} d_i - \Delta}. \quad (8)$$

From Equation 8 follows that the reception duty-cycle is minimized when $\Delta = 0$, and hence $T_C = k \cdot \sum_{i=1}^{n_C} d_i$. The intuition behind Theorem 4 is that if Equation 7 is not satisfied, then $T_C$ can be increased and therefore, the reception duty-cycle $\gamma$ can be reduced without requiring any additional beacons to guarantee discovery with the same $L$. In other words, the coverage intrinsically induced if Equation 7 is not satisfied exceeds what is needed for determinism. Equation 7 is a necessary, but not a sufficient condition for optimal discovery latencies. When Equation 7 is fulfilled, a beacon sequence $B'$ that leads to disjoint coverage exists, but a different one might still induce redundancy. By combining Theorem 3 and 4, we can derive a bound for unidirectional beacons.

**Theorem 5 (Fundamental Bound for Unidirectional Beaconing):** Given a device $E$ that runs an infinite beacon sequence $B_{E, \infty}$ with a duty-cycle of $\beta_E$ and a device $F$ that runs an infinite reception window sequence $C_{F, \infty}$ with a duty-cycle of $\gamma_F$, the minimum worst-case latency that can be guaranteed for $F$ discovering $E$ is as follows.

$$L = \left\lceil \frac{1}{\gamma_F} \cdot \frac{\omega}{\beta_E} \right\rceil. \quad (9)$$

Clearly, optimal values of $\gamma_F$ are of the form $i/k$, $k \in \mathbb{N}$ and other values of $\gamma_F$ do not lead to an improved $L$. 

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Proof: By combining $T_C = k \cdot \sum_{i=1}^{n} d_i$ from Theorem 4 and Equation 1, we can write Equation 6 as follows.

\[ L = \frac{T_C}{\sum_{i=1}^{n} d_i} \cdot \frac{\omega}{\beta_E \cdot \gamma^F}. \] (10)

This holds true for $\gamma^F$ in the form of $1/k$, $k \in \mathbb{N}$. The proof for other duty-cycles follows from the above discussion. \(\square\)

\section*{Mean Latencies:}

This also allows us to compute the mean latency of an optimal ND protocol. In an optimal deterministic protocol, the number of beacons transmitted until discovery, $k$, always lies within $[1, M]$, and each value of $k$ occurs with a probability of $1/M$. In other words, the range of possible initial offsets is subdivided into $M$ partitions, into which the first beacon in range might fall. If it falls into partition $0 \leq k \leq M$, then exactly $k$ beacons need to be sent until some beacon coincides with a reception window. Measured from the first beacon transmitted in range, the expected packet-to-packet discovery latency $E(l^*)$ is hence $\sum_{k=0}^{M} \frac{1}{M} \cdot k \cdot \lambda$, where $\lambda$ is determined by the reception duty-cycle $\beta$. With $M = \frac{1}{\beta}$ (for values of $\gamma$ for which $1/\gamma$ is an integer), this leads to a beacon-to-bacon latency of $E(l^*) = \frac{1}{2}(1/\beta - 1) \cdot \omega^*/\beta$. The additional latency from the point in time at which both devices come into the reception range until the first beacon is transmitted is given by $1/2 \cdot \lambda$, leading to the following expected mean latency.

Lemma 3: For duty-cycles $\gamma$ for which $1/\gamma$ is an integer, any unidirectional ND protocol that provides optimal worst-case latencies, as given by Equation 9, provides the following mean discovery latency.

\[ E(l) = \frac{1}{2} \cdot \frac{\omega}{\gamma^\beta}. \] (11)

It is intuitive that in the worst-case, the first beacon in range falls into the partition leading to a latency of $M \cdot \lambda$, whereas in the average case, a latency of $M/2 \cdot \lambda$ is induced. Therefore, the mean latency from Equation 11 is exactly half of the worst-case latency from Equation 9. Equation 11 represents the lowest mean latency any unidirectional protocol that provides optimal worst-case latencies can provide. It is not clear whether a protocol that does not necessarily provide optimal worst-case latencies can provide even lower mean latencies. In Section VI-C, we compare Equation 11 to the mean latencies achieved by probabilistic protocols.

\section*{B. Symmetric ND Protocols}

In this section, we extend Theorem 5 towards bidirectional (i.e., device $E$ discovers device $F$ and vice-versa), symmetric (i.e., both devices $E$ and $F$ have the same power budget $\eta$) ND. For achieving bidirectional discovery, every device runs both a beacon and a reception window sequence. We in this paper refer to symmetric ND as both devices always running the same tuple $(B_\infty, C_\infty)$. We furthermore assume that $B_\infty$ and $C_\infty$ can be designed such that both sequences on the same device never overlap with each other. We relax this assumption in the appendix.

1) Bi-Directional Discovery: We can achieve bidirectional ND by running the optimal sequences $B_\infty$ and $C_\infty$ we have identified for unidirectional beaconing on both devices simultaneously. The latency of each partial procedure (viz., the discovery of $E$ by $F$ and of $F$ by $E$) is bounded by Theorem 5. As a result, the worst-case latency for both partial discoveries being successful is also bounded by Theorem 5. Since both devices transmit and receive, we optimize the share between $\beta$ and $\gamma$ to obtain the following bound.

\textit{Theorem 6 (Symmetric Bound for Bi-Directional ND Protocols):} For a given power budget $\eta$, no bi-directional symmetric ND protocol (i.e., every device runs the same tuple $(B_\infty, C_\infty)$) can guarantee a lower worst-case latency than the following.

\[ L = \min \left( \frac{2}{\eta} \cdot \frac{\omega^*/\eta - 1}{2}, \frac{2}{\eta} \cdot \frac{\omega^*/\eta - 1}{2} \right). \] (12)

Proof: Theorem 4 implies that optimal values of $\gamma$ are given by $1/\gamma = k, k = 1, 2, 3, \ldots$. Inserting $\eta = \alpha \beta + \gamma$ (cf. Definition 5) into Equation 9 and using $1/\gamma = k$ leads to

\[ L = k^2 \cdot \omega^*/k^{\eta - 1}, k \in \mathbb{N}. \] (13)

We now have to find the value of $k$ that minimizes $L$. Let us for now allow non-integer values of $k$ in Equation 13. By forming the first and second derivative of Equation 13 by $k$, one can show that a local minimum of $L$ exists for $k = \lfloor 2/\eta \rfloor$, which is a non-integer number for most values of $\eta$. By analyzing $dL/dk$, we can further show that Equation 13 is monotonically decreasing for values of $k < \lfloor 2/\eta \rfloor$ and monotonically increasing for values of $k > \lfloor 2/\eta \rfloor$. Hence, the only integer values of $k$ that potentially minimize $L$ are $\lfloor 2/\eta \rfloor$ and $\lceil 2/\eta \rceil$. Inserting $k = \lfloor 2/\eta \rfloor$ or $k = \lceil 2/\eta \rceil$ into Equation 10 and taking the minimum latency among both possibilities leads to Equation 12. \(\square\)

Theorem 6 also holds true for unidirectional beaconing, if the power budget $\eta = \alpha \cdot \beta_E + \gamma_F$ of two devices $E$ and $F$ is to be optimized. Further, one can easily see that for small values of $\eta$, the floor- and ceiling functions in Equation 12 only marginally affect $L$, which can therefore be approximated by

\[ L = \frac{4\alpha \omega}{\eta^2}. \] (14)

Even when both devices $E$ and $F$ transmit as well as receive, it is possible to design unidirectional protocols in which only one of the two devices, $E$ or $F$, can discover the other. Here, the beacons on both devices contribute to a joint notion of coverage, leading to a reduced latency bound compared to the case where both devices can discover each other mutually. A bound for this possibility is given in the appendix.

2) Collision-Constrained Discovery: For achieving the bound given by Theorem 6, we have assumed that the beacons of multiple devices never collide. This assumption is reasonable for a pair of radios, in which collisions only rarely occur. However, as soon as more than two radios are carrying out the ND procedure simultaneously, collisions become inevitable and some of the discovery attempts fail. As a result, some devices might discover each other after the theoretical worst-case latency has passed, or, depending on the protocol design, might not discover each other at all. Therefore, it is often required to limit the channel utilization and hence collision rate, which leads to an increased worst-case latency bound. In protocols with disjoint sequences (i.e., every $\Phi_1$ is covered exactly once), every collision will lead to a failure.
of discovering within $L$. The collision probability is solely
determined by the channel utilization $\beta$. We in this section
study the worst-case latency that can be achieved if both $\eta$
and $\beta$ (and hence the collision probability) are given. We in
addition discuss possibilities to reduce the number of failed
discoveries for a given collision probability in Section VII-B.1.
Consider a number of $S$ senders, of which each occupies
the channel by a time-fraction of $\beta$. The first beacon of an
additional sender that comes into range at any random point
in time will face a collision probability of (cf. [55]):

$$P_c = 1 - e^{-2(1-S)\beta}. \quad (15)$$

Once a beacon has collided, the repetitiveness of infinite bea-
con sequences (cf. Lemma 2) implies that the fraction of later bea-
cons colliding with this device is predefined. Nevertheless,
since all offsets between the two sequences occur with the
same probability, the collision probability of every individual
beacon is given by Equation 15. Note that this equation implies
that the collision probability is independent of the reception
duty-cycle $\gamma$. When constraining the channel utilization to
a maximum value $\beta_m$ that must never be exceeded, the
following latency bound applies.

**Theorem 7 (Bound for Symmetric ND with Constrained
Channel Utilization):** For a given upper bound on the channel
utilization $\beta_m$, no symmetric ND protocol can guarantee a
lower worst-case latency than the following.

$$\begin{align*}
L &= \begin{cases} 
\min(A, B), & \text{if } \eta \leq \gamma_0 + \alpha \beta_m, \\
\frac{1}{\eta - \alpha \beta_m} \cdot \frac{\omega}{\beta_m}, & \text{if } \eta > \gamma_0 + \alpha \beta_m.
\end{cases} 
\quad (16)
\end{align*}$$

Here, $A$ and $B$ are given by Equation 12 and $\gamma_0 = \frac{1}{2[\gamma_n]}$
if $A \leq B$, and $\frac{1}{\gamma_n}$ otherwise.

**Proof:** Given $\eta$, if the channel utilization that results from
choosing the optimal value of $\gamma$ (see proof of Theorem 6)
does not exceed $\beta_m$, the bound given by Equation 12 remains unchanged. Otherwise, the bound is obtained from Equation 9 by eliminating $\gamma$ using $\eta = \alpha \beta_m + \gamma$ (cf. Definition 5). \hspace{1cm} \Box

**C. Asymmetric Discovery**

So far, we have assumed that two devices $E$ and $F$ run the
same tuple of sequences. Often, different devices have different
energy budgets, which can be due to different capacities or
states-of-charge of their batteries, different energy harvesting
capabilities, or different required lifetimes. In such scenarios,
ND protocols that allow all devices to have different power
budgets are required, and hence the sequences on both devices
differ. Next, we study the latencies of asymmetric protocols with
different sequences on both devices, which allow for
configurations with $\eta_E \neq \eta_F$. We thereby assume that each
device knows the tuple of sequences on its opposite device.
This scenario is relevant e.g., when connecting a gadget with
limited power supply to a smartphone using BLE. Here, different
sequences on both devices, which account for their different power budgets, can be specified. The case of every
device being allowed to choose its power budget autonomously
at runtime, and hence, asymmetric ND in which devices are
unaware of the sequences of remote devices, is also relevant.
The possible degradation of the optimal performance in this
case needs to be studied in further work.

1) **Simplified Bound:** We first consider tuples of power bud-
gets $(\eta_E, \eta_F)$, for which $\frac{1}{2\eta_E}$ and $\frac{1}{2\eta_F}$ are integers. We then
extend this towards all power budgets.

**Theorem 8 (Simplified Bound for Asymmetric ND):** Consi-
der two devices $E$ and $F$ with power budgets $\eta_E$ and $\eta_F$, where
$\frac{1}{2\eta_E}$ and $\frac{1}{2\eta_F}$ are integers. The lowest worst-case latency
for two-way discovery is as follows.

$$L = \frac{4\omega}{\eta_E \cdot \eta_F}. \quad (17)$$

**Proof:** According to Theorem 5, if $\frac{1}{2\gamma_E}$ and $\frac{1}{2\gamma_F}$ are integers,
the lowest worst-case one-way discovery latency $L_F$
for device $F$ discovering device $E$ and the latency $L_E$ for
the reverse direction are as follows.

$$L_F = \frac{\omega}{\gamma_F \cdot \beta_E} \quad \text{and} \quad L_E = \frac{\omega}{\gamma_E \cdot \beta_F}. \quad (18)$$

The global worst-case latency for two-way discovery is given
by $L = \max(L_E, L_F)$. Because of this, every optimal asymmetric
ND protocol must fulfill $L_F = L_E$, since in cases of e.g., $L_F > L_E$, one could decrease the reception
duty-cycle $\gamma_E$ of device $E$. In turn, one could increase $\beta_E$, thereby reducing $L_F$ and hence $L$ for the same $\eta_E$. Since
Equation 18 is continuous and differentiable, $L_F = L_E$ can
always be realized. From $L_F = L_F$ and Equation 18 follows that
$\gamma_E / \gamma_F = \beta_E / \beta_F = \text{const} = \mu$. By substituting $\beta_E$ by $\beta_f / \mu$
in $L_F$ (cf. Equation 18) and by substituting $\gamma_F = \gamma_F - \alpha \beta_F$, we obtain $L_F = \frac{\omega}{\gamma_F - \alpha \beta_F}$. By differentiating $L_F$ by $\beta_F$, we can show that $L_F$ is minimal for $\beta_F = \eta_F / 2$ and hence $\gamma_F = \eta_F / 2$. Similarly, $L_E$ has a local minimum at $\beta_E = \eta_E / 2$.

Here, $\gamma_F / \gamma_E$ and $\beta_F / \beta_E$ are integers, also $\gamma_E / \gamma_F$ and $\beta_E / \beta_F$ are integers, and hence Equation 18 holds true. When re-substituting $\mu$ by $\beta_E / \beta_F$ and replacing $\beta_E$ and $\beta_F$ by their optimal values, we obtain Equation 18. \hspace{1cm} \Box

2) **Generic Asymmetric Bound:** For arbitrary power bud-
gets, the latency $L_E$ for device $E$ discovering $F$ and $L_F$
for the reverse discovery result from Equation 9 as follows.

$$L_E = \frac{1}{\gamma_E} \frac{\omega \alpha}{\eta_F - \gamma_E}, \quad L_F = \frac{1}{\gamma_F} \frac{\omega \alpha}{\eta_E - \gamma_F}. \quad (19)$$

We know from Theorem 4 that only values of $\gamma_E$ and $\gamma_F$, for
which $\frac{1}{2\gamma_E}$ and $\frac{1}{2\gamma_F}$ are integers potentially minimize $L_E$
and $L_F$. This also becomes evident from Equation 19. If e.g.,
$\frac{1}{2\gamma_E}$ exceeds its next-lower integer value, we could decrease
$\gamma_E$ and therefore decrease $L_E$. Because only certain discrete
values of $\gamma_E$ and $\gamma_F$ are optimal, the latency functions become
discontinuous and hence, $L_E = L_F$ cannot always be realized.
Finding the tuple of integers that minimizes $L = \min(L_E, L_F)$
is not straightforward, since the number of integers is infinite
and the optimal solution cannot be found using analytical
methods. We in the following first give a solution using
which the second integer of the tuple that minimizes $L$ can
be computed using analytical methods, if the first one is
given. Next, we present an algorithm that limits the solution
space of the remaining integer to a finite number of integers.
By iterating through the resulting candidate solutions, the
configuration that minimizes $L$ can be identified with low
computational complexity. Towards this, it is beneficial to re-
write Equation 19 as follows.
where $L_E^* = \left\lfloor \frac{1}{\gamma_E} \cdot \frac{\omega k}{\eta_E} \right\rfloor \cdot \left(\frac{k \omega \alpha}{\eta_E} \right) - 1 - 1$, $L_r = \frac{k \omega \alpha \Delta E}{\eta_F} - 1$.

Proof: All values of $\Delta E$ for which $\gamma_E$ is an integer are given by the following equation:

$$\Delta E(k) = \frac{1}{k - \frac{\eta E}{2}}, \quad k = \left\lfloor \frac{2}{\gamma E} \right\rfloor + 1, \infty \in \mathbb{N}.$$  \hfill (24)  

Figure 6 depicts $L_E, L_E^*, L_F, L_F^*$ for a given optimal value of $\Delta E$.

values of $\Delta F$ for which $1/\gamma_F$ is an integer will lead to a larger latency of either $L_E$ or $L_F$ (cf. Figure 6). When replacing $\Delta E$ by $\Delta E(k)$ from Equation 24, rounding $1/\gamma_F$ to the next higher integer results into $L_1$, to the next lower one into $L_r$. □

**Optimizing $\eta_F$:** All positive integer values $k$ of $\gamma_F$ are given by Equation 24. Which integer will minimize the worst-case latency? We in the following discuss finding the value of $k$ that minimizes $L_1$. However, the same procedure also holds true for optimizing $L_r$. Differentiating Equation 23 is not possible, since it contains a ceiling term. For this reason, we cannot directly identify its minimum by solving $\frac{dL_1}{dk} = 0$. However, by exploiting the relation $x \leq \left\lfloor x \right\rfloor \leq x + 1$, we can derive a differentiable upper and a lower bound for $L_1$ from Equation 20. These bounds are as follows.

$$L_1^u = \frac{k \omega \alpha}{k \eta_\alpha - 1}, L_1^l = k \cdot \frac{\eta_E}{\eta_F} \cdot \frac{k \omega \alpha}{k \eta_\alpha - 1}.$$  \hfill (25)  

Figure 7 depicts $L_1^u$ and $L_1^l$. It also shows $L_1$, which always lies in-between. By analyzing the derivative of $L_1^u$, one can easily identify the minimum of $L_1^u$, which lies at $k_0$. The value $k_0$ is not necessarily an integer. Clearly, $L_1(k_0)$ will always lie below $L_1^u(k_0)$ (cf. Figure 7). We now solve $L_1^u(k) = L_1^u(k_0)$, and obtain the values $k_1$ and $k_2$, with $k_1 < k_2$. Since $L_1 \leq L_1$, all integer values $k$ that are potentially optimal lie between $[k_1]$ and $[k_2]$ (cf. Figure 7). Note that $k_1$ and $k_2$ always exist, since there is exactly one minimum of $L_1^u$ and $L_1^u$, and $L_1^u < L_1^u$. Further, no other values of $k$ can be optimal, since no value of $L_1$ can be smaller than the corresponding value of $L_1^u$, and all values of $L_1^u$ that lie outside of $\left\lfloor [k_1], [k_2] \right\rfloor$ always exceed those that lie within $\left\lfloor [k_1], [k_2] \right\rfloor$.

With the above, the following scheme is guaranteed to result in the lower bound $L(\eta E, \eta F)$ for asymmetric discovery within a finite number of computational operations.

1. Compute $k_1$ and $k_2$ by solving $L_1^u = L_1(k_0)$
2. Compute the minimum latency $L_{1,\text{min}}$ by evaluating Equation 23 for all values of $k \in \left\lfloor [k_1], [k_2] \right\rfloor$.
3. Repeat Steps 1) and 2) for $L_r$, which leads to $L_{r,\text{min}}$.
4. The worst-case latency $L(\eta E, \eta F)$ is given by $\min(L_{1,\text{min}}, L_{r,\text{min}})$.

The required computation time is negligible, allowing for the computation of $L$ for large numbers of different power budgets within milliseconds on a laptop. Figure 8 exemplifies this bound for asymmetric ND. Here, $\eta E = 0.5$, while we sweep through all values of $\eta F$. Values for which the simplified bound from Theorem 8 applies are highlighted using a circle.
VI. PREVIOUSLY KNOWN PROTOCOLS

In this section, we relate the worst-case performance of popular protocols and previously known bounds to the fundamental limits described in the previous section. We thereby consider symmetric bi-directional discovery. Due to their relevance in practice, we consider only small power budgets $\eta$. For such budgets, the numerical difference between the simplified bound for symmetric protocols given by Equation 14 and the exact bound given by Equation 12 is negligible, allowing for a simplified presentation. We in this section distinguish between slotted and slotless protocols. Recall that slotted protocols are characterized by temporally coupling reception and transmission into the same slot, whereas slotless protocols decouple them from each other. In the schedules of slotted and slotless protocols, time might either be regarded as continuous or discretized. Known optimal protocols, e.g., [18], actually support describing their schedules using discretized time.

A. Worst-Case Bound of Slotted Protocols

As already described in Section II, a worst-case number of slots within which discovery can be guaranteed is known for slotted protocols [20], [21]. The corresponding worst-case latency in terms of time is proportional to the slot length $I$, for which there is no known lower limit. In this section, we transform this worst-case number of slots into a latency bound and establish the relations to the fundamental bounds on ND presented in this paper. We will also address the bound for symmetric protocols given by Equation 14 and [11] is lower in terms of slots than the bound in [20] and [21]. It is achieved by assuming two beacon transmissions per active slot ( [20], [21] assumes only one), of which one beacon is sent slightly outside of the slot boundaries. By accounting for the two beacons per active slot, Equation 26 becomes $\eta = \frac{k \cdot (I + \alpha \omega)}{L}$, which leads to the following bound for the protocols proposed in [10] and [11]:

$$L \geq \frac{(1 + 2\alpha + \alpha^2)}{\eta^2}.$$  

For $\alpha = 1$, this bound becomes $\frac{I}{\eta^2}$ and hence identical to the fundamental bound for symmetric protocols given by Theorem 6. For all other values of $\alpha$, this bound exceeds the one given by Theorem 6. However, the assumption of full-duplex radios is not fulfilled by most wireless devices. Further, every wireless radio requires a time to switch from transmission to reception, called turnaround time, during which the radio is unable to receive any beacons. Even for recent radios, this time is large against the beacon transmission duration $\omega$ (e.g., for the nRF52840 radio, it lies around 40 $\mu$s [56], whereas beacons can be as short as 32 $\mu$s). Therefore, $I$ will be orders of magnitude larger than $\omega$, which linearly increases the worst-case latency of slotted protocols can guarantee in practice.

We now study the bound presented in [10] and [11], which has been claimed to be lower in terms of slots than the one presented in [20] and [21]. It is achieved by assuming two beacon transmissions per active slot ( [20], [21] assumes only one), of which one beacon is sent slightly outside of the slot boundaries. By analyzing the two beacons per active slot, Equation 26 becomes $\eta = \frac{k \cdot (I + 2\alpha \omega)}{L}$, which leads to the following bound for the protocols proposed in [10] and [11]:

$$L \geq \frac{(1 + 2\alpha + \alpha^2)}{\eta^2}.$$  

This bound becomes minimal for $\alpha = 1/2$, for which it is identical to the bound in Theorem 6. Hence, the bound in [10] and [11] is lower in terms of slots than the bound in [20] and [21], but identical or larger in terms of time.

2) Latency/Power Budget/Channel Utilization Metric: All previously known bounds for slotted protocols are in the form of relations between the worst-case number of slots and the power budget or duty-cycle. The channel utilization, which is directly related to the beacon collision rate, has not been considered before. However, in slotted protocols, the channel utilization depends both on the number of active slots per period and on the slot length. For sufficiently large slot lengths, the turnaround times of the radio only play a negligible role. Further, the time for reception in each slot approaches nearly
the whole slot length $I$. Hence, for $I \gg \omega$, we can compute the power budget of slotted protocols as follows.

$$\beta = \frac{k \omega}{IT}, \quad \gamma = \frac{k I}{IT} = \frac{k}{T}, \quad \eta = \gamma + \alpha \beta. \quad (29)$$

With the requirement of $k \geq \sqrt{T}$ from [20] and [21], one can express the slot length $I$ by the desired channel utilization $\beta$ in Equation 29, which results in the following bound.

$$L = \frac{\omega}{\eta \beta - \alpha \beta^2}. \quad (30)$$

From comparing Theorem 7 (cf. Equation 16) to Equation 30, it follows that if $\beta_m$ lies below $\eta/2\alpha$, the worst-case latency of a slotted protocol can guarantee with a channel-utilization of $\beta = \beta_m$ is identical to the corresponding fundamental bound. For $\beta_m > \eta/2\alpha$, slotted protocols cannot reach the fundamental bound from Theorem 7. Figure 10 visualizes this fundamental bound and the bound for slotted protocols from Equation 30. As can be seen, they coincide for low channel utilizations $\beta$, but the worst-case latency of slotted protocols is increased for higher channel utilizations. This implies that slotted protocols can potentially provide optimal worst-case latencies when no device is allowed to utilize the channel beyond a certain threshold, e.g., for limiting the interference with neighboring networks. But when no such constraints exist, they cannot perform optimally.

We in the following evaluate the popular protocols Disco [7], Searchlight-Striped [9], U-Connect [8], Quorum [6], [57], and diffcode-based protocols [20] and compare them to the performance bound given by Theorem 7. In Disco, active slots are repeated every $p_1$ and $p_2$ slots, where $p_1$ and $p_2$ are coprimal numbers. The Chinese Remainder Theorem implies that there is a pair of overlapping slots among two devices every $p_1 \cdot p_2$ time-units. U-Connect, for which we here assume the slot design used in Disco, also relies on coprimal numbers for achieving determinism. In contrast, Searchlight defines a period of $T$ and a hyper-period of $T^2$ slots. The first slot of each period is active, whereas a second active slot per period systematically changes its position, until all possible positions have been probed. Diffcode-based solutions are built on the theory of block designs and hence guarantee a pair of overlapping slots among two devices with the minimum possible number of active slots per worst-case latency. Additional details on the studied protocols can be found in [58].

Slot length-dependent equations on the worst-case latency and duty-cycle of these protocols (assuming an equal power consumption for transmission and reception) are available from the literature. We here scale the slot length and protocol-specific parameters such that a given tuple $(\eta, \beta)$ is realized. In this way, equations on the worst-case latency from the literature, which are given in terms of slots, are transformed into equations in terms of time. The resulting equations are given in Table I. A detailed derivation of these Equations can be found in Appendix F. In the form in Table I, the latencies can be directly compared with the bound from Equation 16. A simplified version of this Equation is also given in Table I, in which a given channel utilization $\beta$ is enforced instead of assuming a constraint that is not to be exceeded. Clearly, only Diffcode-based schedules reach the optimal performance in this metric, whereas all other ones perform below the optimum. In summary, slotted protocols can perform optimally in the latency/power budget/channel utilization metric, if the channel utilization remains low. In the latency/duty-cycle metric, however, higher required channel utilizations prevent slotted protocols from performing optimally.

### B. Worst-Case Bound of Slotless Protocols

1) Latency/Power Budget Metric: We first discuss the case when the channel utilization of every device is not limited. This is typically the case when only a few devices are in range and hence, collisions play a negligible role. Here, every device can optimize its channel utilization to minimize the latency that is guaranteed for a given power budget. We will study the case of a constrained channel utilization in Section VI-B.2.

In slotted protocols, the number of beacons is always coupled to the number of reception windows. As a result, such protocols cannot optimize their channel utilization and hence lack optimality in the latency/power budget metric. Slotless protocols are not subjected to this constraint. Can they reach optimal latency/power budget relations? In [18], two parametrization schemes for slotted protocols, called SingleInt and MultiInt, have been proposed, which have been claimed to provide the best latency/power budget performance among all known slotless protocols. We therefore in the following relate their performance to the bounds presented in Section V.

In such slotless protocols, beacons are sent periodically with a period $T_B$. Similarly, the device listens to the channel for $d_s$ time-units once per period $T_C$. The SingleInt scheme specifies the following configuration: $T_B = d_s - \omega$, $T_C = (M+1) \cdot T_B$, $M = 1, 2, 3, \ldots$ The values of $d_s$ and $M$ are chosen based on the specified power budget as described below - the interested reader may refer to [18] for details. One can easily verify that

| Protocol          | $L(\beta, \eta)$ |
|-------------------|------------------|
| Diffcodes [20], [21] | $\frac{\omega}{\eta \beta - \alpha \beta^2}$ |
| Disco [7]         | $\frac{\omega}{\eta \beta - \alpha \beta^2}$ |
| Searchlight-S [9] | $\frac{3 \omega + \sqrt{\omega^2 (8 \eta - 8 \alpha \beta + 9)}}{8 \omega \eta - 8 \omega \alpha \beta}$ |
| U-Connect [8]     | $\frac{\omega (1 - \sqrt{\omega^2 - \eta + 1})^2}{\eta (\eta - \alpha \beta)}$ |
| Quorum [6], [57]  | $\frac{\omega}{\eta \beta - \alpha \beta^2}$ |
| Generic Optimum (cf. Eq. 16) | $\frac{\omega}{\eta \beta - \alpha \beta^2}$ |
such parametrizations lead to a disjoint coverage. Since the distance between two consecutive beacons does not exceed the length of the reception window minus one beacon transmission duration, the ND procedure will be successful within $T_C$ time-units. Therefore, the worst-case latency is as follows (cf. [18]).

$$L = (M + 1)(d_s - \omega) + \omega.$$  \hfill (31)

For our bounds, we have assumed that 1) beacons that are sent within the last $\omega$ time-units of each reception window are received successfully and 2) the transmission duration of the successfully received beacon is neglected. When applying these assumptions to the protocol described above, we can set $\omega = 0$ in Equation 31 and obtain $L = (M+1)d_s$. The length of the reception window, $d_s$, is determined by the power budget the protocol should realize. It is $\eta = d_s/T_C + \omega/T_B$, which can be solved by $d_s$ easily. This leads to $L = \frac{2}{\eta} \omega$.

By forming the first and second derivative of $L$, one can find that $M = 2/\eta - 1$ minimizes $L$. Since $M$ needs to be an integer number, we consider the pair of neighboring integers, i.e., $M_1 = \lceil 2/\eta - 1 \rceil$ and $M_2 = \lfloor 2/\eta - 1 \rfloor$, which leads to:

$$L_1 = \left[ \frac{2}{\eta} - 1 \right]^2 \frac{\omega}{\eta} - \frac{1}{\eta}, \quad L_2 = \left[ \frac{2}{\eta} - 1 \right]^2 \frac{\omega}{\eta} - 1.$$  \hfill (32)

We parametrize the protocol using $M_1$, if $L_1 < L_2$, and using $M_2$, otherwise. We thereby obtain the following latency.

$$L = \min \left( \left[ \frac{2}{\eta} - 1 \right]^2 \frac{\omega}{\eta} - \frac{1}{\eta}, \left[ \frac{2}{\eta} - 1 \right]^2 \frac{\omega}{\eta} - 1 \right).$$  \hfill (33)

This is identical to Theorem 6. Hence, under the assumptions described above, a slotless protocol parametrized using the SingleInt scheme is optimal in the latency/power budget metric. Which degradation of the latency bound of the SingleInt scheme do these assumptions imply in practice? Assuming a beacon transmission duration of $\omega = 32$ $\mu$s and a range of power budgets between 0.1% and 100% in steps of 0.1% and $\alpha = 1$, the normalized root mean square error between Equation 12 and the equations from [18] is 1.24%.

2) Latency/Duty-Cycle/Channel Utilization Metric: Slotless protocols parametrized as described in the previous section always use the channel utilization that minimizes $L$. Therefore, they cannot obey a given limit on the channel utilization. In scenarios with a large number of devices in range, the maximum channel utilization might be constrained. This requires a different protocol than those described in the previous section, which always optimize their channel utilization for the lowest possible $L$ given a power-budget $\eta$. Next, we propose a parametrization scheme for PI protocols that can account for a given limit on the channel utilization $\beta_m$ and show that the resulting latencies are optimal for a given power budget and maximally allowed channel utilization $\beta_m$.

Let us again assume $T_B = d_s$ and $T_C = (M + 1)d_s$. Here, the channel utilization is given by $\beta = \omega/d_s$. Hence, $\beta$ can be controlled by the length of the reception window $d_s$. In particular, for enforcing $\beta \leq \beta_m$, we have to ensure that $d_s \geq \omega/\beta_m$. By rearranging $\eta = d_s/T_C + \alpha\omega/T_B$ and inserting $T_B = d_s$ and $T_C = (M + 1)d_s$, we obtain $M = (\eta - \frac{\omega}{\eta^2})^{-1} - 1$. Clearly, the larger $d_s$ becomes, the larger also $M$ becomes. The smallest integer value of $M$ for which $d_s \geq \omega/\beta_m$ (and hence, $\beta \leq \beta_m$) is therefore as follows.

$$M = \left[ \frac{1}{\eta - \alpha \beta_m} - 1 \right].$$  \hfill (34)

With $L = (M + 1)d$ (cf. Section VI-B.1), we obtain:

$$L = \left[ \frac{1}{\eta - \alpha \beta} \right] \frac{\omega}{\beta}.$$  \hfill (35)

Section VI-B.1 describes the value of $M$ that minimizes the worst-case latency if the channel utilization is unconstrained. From $\eta = d_s/T_C + \alpha\omega/T_B$ follows that a certain value of $M$ leads to a channel utilization of $\beta = 1/\alpha(\eta - 1/M + 1)$. If the channel utilization obtained for the optimal $M$ from Section VI-B.1 lies below the limit $\beta_m$, we can safely use this value and obtain the worst-case latency given by Equation 33. Otherwise, we have to use the value for $M$ given by Equation 34, leading to the latency given by Equation 35. For all of these cases, the latencies achieved are equal to those given by Theorem 7. Hence, a periodic interval protocol parametrized as described above is the first one to cover the entire Pareto-front given by the power budget, channel utilization and worst-case latency.

Equation 35 describes the lowest discovery latency any slotless protocol can achieve for a given tuple $(\eta, \beta)$. For small values of $\eta$ and $\beta$, the ceiling function can be omitted and we obtain $L = \frac{\omega}{\eta \beta}$. It is easy to verify that this corresponds to the generic optimum given in the last row of Table I.

C. Probabilistic Protocols

It is frequently assumed that probabilistic protocols could potentially provide shorter mean latencies than deterministic ones. We next debunk this assumption. In a probabilistic protocol, the probability that any considered point in time lies within a reception window is given by $\gamma$. Hence, also the probability that any transmitted beacon falls into a reception window is equal to $\gamma$. Therefore, the expected packet-to-packet latency is here:

$$E(l^*) = \sum_{i=0}^{\infty} (1 - \gamma)^i \cdot \gamma \cdot i \cdot \bar{X} = \frac{\omega}{\beta} \cdot (1/\gamma - 1).$$  \hfill (36)

Recall from Section V-A.1 that this equals the worst-case beacon-to-beacon latency of an optimal deterministic protocol. In other words, a probabilistic protocol takes as long on the average as a deterministic protocol in the worst-case. The additional latency from the point in time at which both devices come into range until the first beacon is transmitted is given by $1/2 \cdot \bar{X}$, leading to a mean latency of $\frac{\omega}{\beta} = \frac{1}{\gamma} - 1$. For a probabilistic protocol. Comparing this to Equation 11 shows that the average latency of a probabilistic protocol is approximately $2\times$ that of a deterministic one.

VII. GENERALIZATIONS AND OPEN PROBLEMS

A. Generalizations

The bounds presented in this paper were presented based on multiple assumptions, which can all be relaxed. Such
generalizations are studied in the Appendix, which is available as supplementary material. In Appendix A, we study how these bounds change, if the transmission of the successfully received beacon contributes to the discovery latency. We furthermore take into account that beacons sent within the last $\omega$ time-units of a reception window cannot be received successfully, and study the impact of switching overheads. We consider non-repetitive reception window sequences in Appendix C. The inability of a radio to transmit and receive at the same time is studied in Appendix D, and of clock skew in Appendix B. An additional scenario, in which either of two devices can discover its opposite, is presented in Appendix E.

B. Open Problems

We next discuss the most important open problems.

1) Open Problems On Fundamental Limits: There are two most important open problems related to fundamental limits. First, what is the lowest latency an asymmetric protocol can guarantee, if the power budgets and hence, duty-cycles of all devices are unknown? Second, the bounds derived so far are valid for a pair of devices discovering each other. For unidirectional beaconing, protocols in which 100% of all discovery attempts are successful within $L$ time-units can be realized in practice. Similarly, bidirectional protocols only fail with low probabilities. But for increasing numbers of devices participating simultaneously in the ND procedure, it is inevitable that their beacons collide and hence, an increasing number of discovery attempts fail. Therefore, generalized performance bounds for multi-device scenarios need to be derived. Such bounds are in the form of a function $L(\beta, \gamma, S, P_f)$, which needs to be interpreted as follows. For a given number of senders $S$ with duty-cycles of $\beta$, and a receiver with a duty-cycle of $\gamma$, in no ND protocol, a fraction of at least $1 - P_f$ of all discovery attempts will terminate successfully within less than $L$ time-units. Clearly, for $S \rightarrow 1$ and $P_f \rightarrow 0$, this bound converges to $L$ from Equation 9. The following mechanisms determine the performance in scenarios with $S > 1$.

1) Lowering the Channel Utilization: The rate of collisions is determined by the channel utilization $\beta$, as given by Equation 15. Hence, devices can reduce the failure probability $P_f$ by reducing $\beta$, which will, however, negatively affect the achieved discovery latencies (cf. Equation 9).

2) Redundant Coverage: Optimality in the $L(\beta, \gamma)$ - metric for two devices implies that every initial offset is covered exactly once (cf. Theorems 2 and 4) and hence, every collision for two devices implies that every initial offset is covered. Hence, for such offsets, more than one beacon would collide with a reception window, and as long as one of them does not collide, the discovery procedure would succeed.

The collision of a pair of beacons from two devices often induces an increased collision probability of subsequent pairs of beacons. For example, consider protocols in which beacons are sent with periodic intervals. Since all devices in a symmetric scenario transmit with the same interval, a collision implies that all later beacons will also collide. To make protocols robust against failures due to collisions, a beacon schedule needs to decorrelate the collision probabilities of redundantly coinciding beacons from each other. It is currently not clear which degree of such a decorrelation can be actually achieved. Further, measures for decorrelating collision probabilities might reduce the latency performance, because they could prevent beacons from being sent at their optimal points in time. Therefore, not all initial offsets could be covered using the lowest possible number of beacons, making additional beacon transmissions necessary. Besides open questions on decorrelating collisions, for protocols being optimal in the multiple-device case, how many times should every initial offset be covered? These questions need to be answered in future research.

2) Open Problems on Protocol Design: Our results also outline an important direction for the development of future ND protocols. Protocols that contain decorrelation mechanisms to make the collision of each beacon independent from the occurrence of previous collisions have not been sufficiently studied. Though BLE applies a random delay for scheduling its beacons [48], the optimal decorrelation technique to obtain the best trade-off between robustness and worst-case latency remains an open question.

VIII. CONCLUDING REMARKS

We have presented and proven the correctness of multiple fundamental bounds on the performance of ND protocols. In particular, we have presented bounds for unidirectional beaconing, for symmetric and for asymmetric bi-directional ND. Further, we have shown that in the latency/power budget metric, only slotless protocols can perform optimally. However, if the channel utilization is constrained, slotted protocols can cover large parts of the Pareto-Front, while we have presented a slotless protocol that can cover the entire one.

REFERENCES

[1] M. J. McGlynn and S. A. Borbash, “Birthday protocols for low energy deployment and flexible neighbor discovery in ad hoc wireless networks,” in Proc. 2nd ACM Int. Symp. Mobile ad hoc Netw. Comput.-MobiHoc, 2001, p. 137.
[2] R. Margolies, G. Grebta, T. Chen, D. Rubenstein, and G. Zussman, “Panda: Neighbor discovery on a power harvesting budget,” IEEE J. Sel. Areas Commun., vol. 34, no. 12, pp. 3606–3619, Dec. 2016.
[3] L. You, Z. Yuan, P. Yang, and G. Chen, “ALOHA-like neighbor discovery in low-duty-cycle wireless sensor networks,” in Proc. IEEE Wireless Commun. Netw. Conf., Mar. 2011, pp. 749–754.
[4] S. Vasudevan, D. Towsley, D. Goeckel, and R. Khalili, “Neighbor discovery in wireless networks and the coupon collector’s problem,” in Proc. 15th Annu. Int. Conf. Mobile Comput. Netw., Sep. 2009, pp. 181–192.
[5] C. Schurgers, V. Tsiatsis, S. Ganeriwal, and M. Srivastava, “Optimizing sensor networks in the energy-latency-density design space,” IEEE Trans. Mobile Comput., vol. 1, no. 1, pp. 70–80, Jan. 2002.
[6] Y. C. Tseng, C.-S. Hsu, and T.-Y. Hsieh, “Power-saving protocols for IEEE 802.11-based multi-hop ad hoc networks,” in Proc. 21st Annu. Joint Conf. IEEE Comput. Commun. Soc. Conf., vol. 1, Jun. 2002, pp. 200–209.
[7] P. Dutta and D. Culler, “Practical asynchronous neighbor discovery and rendezvous for mobile sensing applications,” in Proc. 6th Int. Conf. Embedded Netw. Sens. Syst. (SenSys), 2008, pp. 71–84.
[8] A. Kandhalu, K. Lakshmanan, and R. R. Rajkumar, “U-connect: A low-latency energy-efficient asynchronous neighbor discovery protocol,” in Proc. Int. Conf. Inf. Process. Sens. Netw. (IFIPSN), 2010, pp. 350–361.
[9] M. Bakht, M. Trower, and R. H. Kravets, “Searchlight: Won’t you be my neighbor?” in Proc. 18th Annu. Int. Conf. Mobile Comput. Netw., Aug. 2012, pp. 185–196.
[10] T. Meng, F. Wu, and G. Chen, “On designing neighbor discovery protocols: A code-based approach,” in Proc. IEEE Conf. Comput. Commun. (IEEE INFOCOM), Apr. 2014, pp. 1689–1697.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Y. Qiu, S. Li, X. Xu, and Z. Li, “Talk more listen less: Energy-efficient
neighbor discovery,” in Proc. IEEE Conf. Comput. Commun. (INFOCOM), Apr. 2015, pp. 693–701.

D. Zhang et al., “ACC: Generic on-demand accelerations for neighbor
discovery in mobile applications,” in Proc. 10th ACM Conf. Embedded Netw. Sensor Syst., Nov. 2012, pp. 169–182.

W. Sun, Z. Yang, K. Wang, and Y. Liu, “Hello: A generic flexible
protocol for neighbor discovery,” in Proc. IEEE Conf. Comput. Commun. (INFOCOM), Apr. 2014, pp. 540–548.

Y. Qiu, S. Li, X. Xu, and Z. Li, “Talk more listen less: Energy-efficient
neighbor discovery in wireless sensor networks,” in Proc. IEEE Conf. Comput. Commun. (INFOCOM), Apr. 2015, pp. 1–9.

C. Julien, C. Liu, A. L. Murphy, and G. P. Picco, “BLEnd: Practical
continuous neighbor discovery protocol for Bluetooth low energy,” in Proc. ACM/IEEE Int. Conf. Infor. Process. Sens. Netw. (IPSN), Jul. 2017, pp. 105–116.

P. Kindt, D. Yunge, G. Reinerth, and S. Chakraborty, “Griassdi: Mutually
assisted slotted neighbor discovery,” in Proc. ACM/IEEE Int. Conf. Infor. Process. Sens. Netw. (IPSN), Jul. 2017, pp. 93–104.

P. H. Kindt, S. Narayanaswamy, M. Saur, and S. Chakraborty, “Optimizing
BLE-like neighbor discovery,” IEEE Trans. Mobile Comput., vol. 21, no. 5, pp. 1779–1797, May 2022.

L. Wei et al., “Lightning: A high-efficient neighbor discovery protocol
for low duty cycle WSNs,” IEEE Commun. Lett., vol. 20, no. 5, pp. 966–969, May 2016.

R. Zheng, J. C. Hou, and L. Sha, “Asynchronous wakeup for ad hoc
networks,” in Proc. 4th ACM Int. Symp. Mobile ad hoc Netw. Comput., Jun. 2003, pp. 35–45.

R. Zheng, J. C. Hou, and L. Sha, “Optimal block design for asyn-
chronous wake-up schedules and its applications in multipath wireless
networks,” IEEE Trans. Mobile Comput., vol. 5, no. 9, pp. 1228–1241, Sep. 2006.

T. Chen, J. Ghaderi, D. Rubenstein, and G. Zussman, “Maximizing
broadcast throughput under ultra-low-power constraints,” IEEE/ACM Trans. Netw., vol. 26, no. 2, pp. 779–792, Apr. 2018.

A. Kandhalu, A. E. Xafa, and S. Hosur, “Towards bounded-latency
Bluetooth low energy for in-vehicle network cable replacement,” in Proc. Int. Conf. Connected Vehicles Expo (ICCVE), Dec. 2013, pp. 635–640.

W. Lee, J.-H. Youn, and T. Song, “Prime-number-assisted block-based
neighbor discovery protocol in wireless sensor networks,” Int. J. Distrib. Sensor Netw., vol. 15, no. 1, Jan. 2019, Art. no. 150514771982624.

K. Kim, H. Roh, W. Lee, S. Lee, and D. Du, “PND: A p-persistent
neighbor discovery protocol in wireless networks,” Wireless Commun. Mobile Comput., vol. 13, no. 7, pp. 650–662, May 2013.

S. Vasudevan, M. Adler, D. Goeckel, and D. Towsley, “Efficient algo-
rithms for neighbor discovery in wireless networks,” IEEE/ACM Trans. Netw., vol. 21, no. 1, pp. 69–83, Feb. 2013.

A. Purohit, B. Priyantna, and J. Liu, “WiFlock: Collaborative group
discovery and maintenance in mobile sensor networks,” in Proc. 10th ACM/IEEE Int. Conf. Infor. Process. Sensor Netw., Apr. 2011, pp. 37–48.

S. Vasudevan, J. F. Kurose, and D. F. Towsley, “On neighbor discovery
in wireless networks with directional antennas,” in Proc. IEEE Conf. Comput. Commun. (INFOCOM), vol. 4, Jul. 2005, pp. 2502–2512.

G. Jakllari, W. Luo, and S. Krishnamurthy, “An integrated neighbor
discovery and MAC protocol for ad hoc networks using directional antennas,” IEEE Trans. Wireless Commun., vol. 6, no. 5, pp. 1024–1114, Mar. 2007.

Z. Zhang and B. Li, “Neighbor discovery in mobile ad hoc self-
configuring networks with directional antennas: Algorithms and com-
parisons,” IEEE Trans. Wireless Commun., vol. 7, no. 5, pp. 1540–1549, May 2008.

K. Twarowski, A. C. Viana, and A. Wolisz, “Optimized asyn-
chronous multi-channel neighbor discovery,” in Proc. IEEE INFOCOM, Apr. 2011, pp. 536–540.

W. Zeng, S. Vasudevan, X. Chen, B. Wang, A. Russell, and W. Wei,
“Neighbor discovery in wireless networks with multipacket reception,” in Proc. 12th ACM Int. Symp. Mobile Ad Hoc Netw. Comput., May 2011, pp. 1–10.

S. A. Borbash, A. Ephremides, and M. J. McGlynn, “An asynchronous
neighbor discovery algorithm for wireless sensor networks,” Ad Hoc Netw., vol. 5, no. 7, pp. 998–1016, Sep. 2007.