

Random Walks with Multiple Step Lengths

Lucas Boczkowski, Brieuc Guinard, Amos Korman, Zvi Lotker, Marc Renault

▶ **To cite this version:**

Lucas Boczkowski, Brieuc Guinard, Amos Korman, Zvi Lotker, Marc Renault. Random Walks with Multiple Step Lengths. LATIN 2018: Theoretical Informatics - 13th Latin American Symposium, Apr 2018, Buenos Aires, Argentina. pp.174-186. hal-01955582

HAL Id: hal-01955582
https://hal.science/hal-01955582

Submitted on 17 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Random Walks with Multiple Step Lengths *

Lucas Boczkowski[1], Brieuc Guinard[1], Amos Korman[1], Zvi Lotker[2], and Marc Renault[3]

[1] IRIF, CNRS and University Paris Diderot, Paris, France.
[2] Ben Gurion University of the Negev, Beersheba, Israel
[3] Computer Sciences Department, University of Wisconsin - Madison.

Abstract. In nature, search processes that use randomly oriented steps of different lengths have been observed at both the microscopic and the macroscopic scales. Physicists have analyzed in depth two such processes on grid topologies: *Intermittent Search*, which uses two step lengths, and *Lévy Walk*, which uses many. Taking a computational perspective, this paper considers the number of distinct step lengths k as a *complexity measure* of the considered process. Our goal is to understand what is the optimal achievable time needed to cover the whole terrain, for any given value of k. Attention is restricted to dimension one, since on higher dimensions, the simple random walk already displays a quasi linear cover time.

We say X is a *k-intermittent search* on the one dimensional n-node cycle if there exists a probability distribution $\mathbf{p} = (p_i)_{i=1}^{k}$, and integers $L_1, L_2, \ldots, L_k$, such that on each step X makes a jump $\pm L_i$ with probability p_i, where the direction of the jump (+ or −) is chosen independently with probability 1/2. When performing a jump of length L_i, the process consumes time L_i, and is only considered to visit the last point reached by the jump (and not any other intermediate nodes). This assumption is consistent with biological evidence, in which entities do not search while moving ballistically.

We provide upper and lower bounds for the cover time achievable by k-intermittent searches for any integer k. In particular, we prove that in order to reduce the cover time $\Theta(n^2)$ of a simple random walk to linear in n up to logarithmic factors, roughly $\frac{\log n}{\log \log n}$ step lengths are both necessary and sufficient, and we provide an example where the lengths form an exponential sequence.

In addition, inspired by the notion of intermittent search, we introduce the *Walk or Probe* problem, which can be defined with respect to arbitrary graphs. Here, it is assumed that querying (probing) a node takes significantly more time than moving to a random neighbor. Hence, to efficiently probe all nodes, the goal is to balance the time spent walking randomly and the time spent probing. We provide preliminary results for connected graphs and regular graphs.

* This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 648032).

1 Introduction

1.1 Background and Motivation

The theory of random walks was first studied in an attempt to abstract the movement of natural entities, such as particles or insects [16, 34]. The term "random walk" itself was originally coined by Karl Pearson in 1905 aiming to model the movement of a mosquito in a forest [34]. Random walk theory has since attracted the attention of researchers in many different disciplines, and has developed into one of the most impressive manifestations of a cross-disciplinary theory. In particular, motivated by the simplicity of this memoryless algorithm, mathematicians and computer scientists have studied random walks intensively, focusing mostly on analyzing its time complexities in finite graphs [2, 12, 17, 18, 28, 29], as well as on identifying various applications of it in different, often seemingly unrelated, computational contexts, see e.g. [3, 22, 35] among many others.

In the last couple of decades, empirical evidence has suggested that in various natural contexts, movement appears to be similar to random walks, yet with heterogeneous step lengths (jumps). Examples appear both on the microscopic scale, such as in the reaction pathway of DNA binding proteins, immune cells movement, crawling amoeba, optics, and in low dimension Hamiltonian chaos [8, 13, 21, 25], as well as on the macroscopic scale, such as in albatrosses, bumblebees, deer, and even humans [4, 5, 9, 26, 36, 37, 39, 40][†]. Most of these examples appear in search contexts, e.g., searching for pathogens or food. It has been further argued in these works that in the corresponding contexts, biological entities alternate between slow diffusing phases in which targets can be detected, and faster phases of ballistic movements (which are typically more rare) during which the search efficiency is weak, effectively allowing targets to be found only between jumps. This compromise between moving and searching has also been studied in deterministic settings [14]. From a search efficiency perspective, it has further been argued that such processes can help to strike a proper balance between global exploration and local exploitation.

Within this family of search strategies, two extreme cases have been extensively studied, namely, *Intermittent search* and *Lévy walks* (see the survey [6] and the references therein). The former process is essentially a random walk with two step lengths: choose an angle uniformly at random (u.a.r), then take a step of unit length with some probability p; otherwise, take a step of some predetermined larger length L. In the latter process, step lengths have a probability distribution that is heavy-tailed: at each step an angle is chosen u.a.r, and the probability to perform a step of length d is proportional to $d^{-\alpha}$, for some fixed parameter $\alpha > 1$. Based on differential equation techniques, these two types of processes have been studied by physicists, aiming to optimize the parameters involved in order to minimize the hitting time under various target

[†] Some of these statistical findings which claim that these animals perform Lévy walks have recently been under debate, due to the difficulty of fitting empirical data to a particular distribution of step lengths [15, 38].

distributions in continuous Euclidean spaces [6, 10, 30, 31, 33, 40]. For example, [31, 33] showed that on the one dimensional n-cycle of length n, an intermittent search with the right choice of parameters can reduce the cover time of a simple random walk from roughly n^2 to roughly $n^{4/3}$. Lévy walks can reduce the cover time substantially further. Indeed, with a little extra work (as appears in the full version), it follows from [1] that a Lévy walk process with parameter $\alpha = 2$ can reduce the cover time to almost linear (up to polylog factors).

This paper studies random walk processes with multiple step lengths from a more unified computational perspective. Specifically, driven by the plausible assumption that utilizing more step lengths (while associating to each a tailored probability) may require more computational resources, our main subject of interest concerns quantifying the trade-off between the number of step lengths and the best possible search performances. Specifically, we are interested in the best cover time achievable by a random walk that uses k step lengths for every integer k. From a technical point of view, the challenge lies in understanding what is the best possible balance between actions on different scales, ranging from highly local ones (small jumps) to highly global ones (large jumps) [24, 27].

The underlying topology we concentrate on is a discrete cycle of n nodes. It is possible to define the notion of random walks with multiple step lengths in tori or grids of all dimensions, but in this preliminary work, we focus on one dimension since this case enjoys the highest increase in performance as the number of step lengths grows. Indeed, simple random walks are already highly competitive in higher dimensions. We note that one dimension search finds relevance in several biological contexts, including in collective navigation by ants [19] and in the reaction pathway of DNA binding proteins [8, 13]. The latter case is a good example of a search with two phases, one which is three-dimensional and fast, the other one-dimensional, slow, which corresponds to sliding along the DNA to find the target site.

Finally, inspired by the notion of intermittent random walks, and aiming to further develop the aforementioned balancing principle, we introduce and investigate a problem, called *Walk or Probe*, which can be defined with respect to arbitrary underlying graphs, and is of independent interest. Here, it is assumed that querying (probing) a node takes significantly more time than moving to a random neighbor. This assumption is consistent with the aforementioned hypothesis that many processes in nature, including e.g., immune cells, cannot engage in moving fast, and, at the same time, execute their search mechanism. Hence, to probe all nodes in a relatively short time, the goal is to balance the time spent walking randomly and the time spent probing. In some sense, a long phase in which the process executes a random walk may be interpreted as "implementing a long jump", in the sense that, with a certain cost, it allows the process to re-start at a different area of the graph.

1.2 Models

k-intermittent search. Let C_n be the n-node cycle and let k be an integer. We say X is a *k-intermittent search* on C_n if there exists a probability distribution $\mathbf{p} = (p_i)_{i=1}^k$, where $\sum_i p_i = 1$, and integers $L_1, L_2, \ldots, L_k$ such that, on each step, X makes a jump $\{0, -L_i, +L_i\}$ with probability respectively $p_i/2, p_i/4, p_i/4$. Overall, with probability $1/2$, the process X stays in place[‡]. The numbers (p_i) and (L_i) are called the *parameters* of the search process X.

Our goal is to show upper and lower bounds on the *cover time* of a k-intermittent search; that is, the expected time to visit every node of the ambient graph C_n, where we assume that a jump from some point to b visits only the endpoint b, and not any of the intermediate nodes. Importantly, we are interested in time rather than the number of moves and, hence, need to account for the travel time of jumps. For simplicity, we assume that the speed of the walker is constant (rather than varying between step lengths), which in particular means that it takes one unit of time to make a move to a neighbor and L units of time to make a jump of length L.

More formally, let us denote by $V_1, V_2, \ldots, V_s, \ldots$ independent random variables taking value L_i with probability p_i for every $i \in [k]$. We also use sign variables $\xi_1, \xi_2, \ldots$ which take value 0 or ± 1 with probability $\frac{1}{2}, \frac{1}{4}, \frac{1}{4}$. We call a product $\xi_s V_s$ a *jump* and ξ_s is the *sign* of the jump. We can then define the move-process $Z(m)$ on $\mathbb{Z}$ and $X(m)$ on the cycle C_n, after m moves, as

$$Z(m) = \sum_{s=1}^{m} \xi_s \cdot V_s, \qquad X(m) = Z(m) \mod n. \tag{1}$$

As we consider it takes one unit of time to travel a distance 1, the time it took to accomplish the first m moves, denoted $T(m)$, is defined as

$$T(m) := \sum_{s=1}^{m} |\xi_s| \cdot V_s. \tag{2}$$

On the finite graph C_n, we denote by M the random number of moves needed before X has visited every node of C_n. The quantity whose expectation we want to bound is $T(M)$, the time needed to visit all nodes, which is called the *cover time*.

Walk or Probe. Consider a simple random walker that walks on a connected graph G and aims to *probe* all nodes in G as fast as possible. The walker at a node is unable to detect whether it has previously probed it. At this point it needs to decide whether to continue the walk or probe it and then continue the walk. Crucially, probing a node is time consuming, and can potentially be very

[‡] This laziness assumption is used for technical reasons, as is common in many other contexts of random walks. However, note with Eq. (2) that this assumption does not affect the time of the process.

slow with respect to the time required to move between neighbors. Specifically, let us assume that each edge traversal costs 1 unit of time, while probing a node costs $C \geq 0$ time units, where C can be a function of several parameters of G (e.g., the number of nodes, edges, or maximal degree).

The Walk or Probe problem aims to find a strategy that balances the time spent in walking vs. probing so as to minimize the *probing cover time*, that is, the expected time until all nodes are probed.

1.3 Our Results

***k*-intermittent search on the cycle.** We report our results in terms of cover time, but the same bounds, divided by a $\log n$ term, apply for hitting times (the time to find any given node), as is clear from our proofs.

Definition 1. *Let B be an integer. Define the* k-intermittent search with base B, *by the parameters*[§] $L_i = B^i$ *for every* $1 \leq i \leq k-1$, $p_i = \frac{1}{L_i}$ *and* $L_0 = 1$ *and* $p_0 = 1 - \sum_{i=1}^{k-1} p_i$.

Theorem 1 *Let k, B, n be integers such that $2 \leq B < n$ and $B^{k-1} \leq n \leq B^k$. The cover time of the k-intermittent search with base B on the n-cycle is at most $poly(k) \cdot poly(B) \cdot n \log n$.*

Hence, from Theorem 1, taking $B = \lceil n^{1/k} \rceil$, we derive the following corollary.

Corollary 2 *For any $k \leq \frac{\log n}{\log \log n}$, there exists a k-intermittent search with cover time $n^{1+O(\frac{1}{k})} \log n$. In particular, if $k = \frac{\log n}{\log \log n}$, then the expected cover time is $n \cdot \log^{O(1)}(n)$.*

Corollary 2 is almost tight, as shown by the following lower bound:

Theorem 3 *For every $\varepsilon > 0$, there exist sufficiently small constants $c, c' > 0$ such that for $k \leq c' \frac{\log n}{\log \log n}$, any k-intermittent search cannot achieve a cover time better than $c \cdot n^{1+\frac{1/2-\varepsilon}{k+1}}$. In particular, for $k = o(\frac{\log n}{\log \log n})$, the cover time is $n \cdot \log^{\omega(1)}(n)$.*

Walk or Probe. Consider the Walk or Probe model on a connected graph $G = (V, E)$ with n nodes, with cost of probing C. Denote by t_{cov} (resp. t_{mix}) the cover time (resp. mixing time) of a random walk on G (see Section 4 for the

[§] Note that $\sum_{i=1}^{k-1} p_i = \sum_{i=1}^{k-1} B^{-i} = \frac{1}{B-1}(1 - \frac{1}{B^{k-1}})$ is a decreasing function of $B \geq 2$, and for $B \geq 3$ is less than $\frac{1}{2}$ so that $p_0 \geq \frac{1}{2}$. For $B = 2$, we would have $p_0 = \Theta(\frac{1}{n})$. In this case, we can change the definition of p_i by dividing by a factor 2, which will allow to have $p_0 \geq \frac{1}{2}$. This does not change anything beyond the constants we use in the proofs and in the results.

definition of the mixing time). The most naive strategy is to probe after each step, in which case we get a probing cover time of:

$$(C+1) \cdot t_{cov}. \tag{3}$$

We show that there exists another simple strategy whose running time depends on t_{mix}, which can be much more efficient in some cases. Let us call *t-strategy* the tactic that consists in probing only once every t steps. Let $G = (V, E)$ be an n-node connected graph.

Theorem 4 *The probing cover time of the $\sqrt{t_{mix}}$-strategy is*

$$O\left((C + \sqrt{t_{mix}}) \cdot |E| \log n\right). \tag{4}$$

And if G is regular, the probing cover time of the $\sqrt{t_{mix}}$-strategy is

$$O\left((C + \sqrt{t_{mix}}) \cdot n \log n\right). \tag{5}$$

In particular, we can derive from Theorem 4 (taking $C = 0$) the following upper bound.

Corollary 5 *For any n-node connected graph $G = (V, E)$,*

$$t_{cov} = \begin{cases} O(\sqrt{t_{mix}} n \log n) \text{ if } G \text{ is regular} \\ O(\sqrt{t_{mix}} |E| \log n) \text{ otherwise.} \end{cases} \tag{6}$$

For the case of regular graphs, the upper bound on t_{cov} mentioned in Corollary 5 was already known ([23, Theorem 1.4]). To the best of our knowledge, the latter bound, for general graphs, is new.

Since the search should cover every node and probe every node of G, we have the following trivial lower bound for the probing cover time:

$$t_{cov} + C \cdot n. \tag{7}$$

Let us compare this lower bound to the upper bounds on the probing cover time of the 1- and $\sqrt{t_{mix}}$-strategies, given by Eq. (3) and (5), in the case of regular graphs.

One family of instances consists of regular graphs for which Eq. (6) is tight (up to poly-logarithmic factors). In this case Eq. (5) becomes, up to poly-logarithmic factors, $Cn + t_{cov}$, and so the $\sqrt{t_{mix}}$-strategy is near optimal. This happens, for example, if the mixing time is poly-logarithmic (e.g. cliques and all expander graphs, such as random r-regular graphs [20]) and in such cases, by Eq. (3), the 1-strategy is also near optimal since the cover time is almost linear as Eq. (6) shows. On the other hand, Eq. (6) is also tight for the path and the cycle, on which the mixing time and cover time are $O(n^2)$. Hence, in this case the $\sqrt{t_{mix}}$-strategy significantly outperforms the trivial 1-strategy when C is large.

If Eq. (6) is not tight then the $\sqrt{t_{mix}}$-strategy may not be optimal. An example where this happens is the torus of dimension $d \geq 2$: Indeed, the mixing time of such a torus is $\Theta(n^{2/d})$, while the cover time is almost linear [29, Sections 5.3.2 and 10.4]. In this case the 1-strategy has probing cover time near optimal by Eq. (3), but the $\sqrt{t_{mix}}$-strategy, which needs at least $\sqrt{t_{mix}}n$ steps (because it must probe at least n times), is not optimal.

Finally, regarding the general bound in Eq. 4 (for not-necessarily regular graphs), let us look at Erdos-Renyi random graphs $\mathcal{G}(n,p)$ with $p = c/n$ for a constant $c > 1$. In this case the mixing time is a.a.s. $\Theta(\log^2 n)$ [7, Theorem 1.1] and the cover time is a.a.s. $\Theta(n \log^2 n)$ [11, Theorem 2.(a)], and the number of vertices is a.a.s. linear in n. Therefore, by Eq. (4), the $\sqrt{t_{mix}}$-strategy has probing cover time $O(Cn \log n + n \log^2 n)$, while the 1-strategy, by Eq. (3), takes time $(C+1)n \log^2 n$, and hence the former strategy gains a logarithmic factor in the number of probes.

1.4 Preliminaries

Notation. When writing logarithms, unless mentioned otherwise, we assume that the base is 2. We denote the cycle of length n by C_n and label its nodes with the serials $\{0, 1, \ldots, n-1\}$. We use the symbols c, c', c'' for constants that we do not compute explicitly. In order to avoid the use of too many symbols, we sometimes employ the same symbol for different constants.

From counting time to counting steps. Recall that $T(m) = \sum_{s=1}^{m} |\xi_s| \cdot V_s$. Let us recall also the well-known

Lemma 6 (Wald's identity). If $(X_t)_{t \geq 0}$ is an i.i.d. sequence with $|E(X_0)| < \infty$, and T is a stopping time for this sequence (i.e. at each time t, the occurrence of the event $T = t$ depends only on $X_0, X_1, \ldots, X_t$), with $|\mathbb{E}(T)| < \infty$, then:

$$\mathbb{E}(\sum_{t=0}^{T} X_t) = \mathbb{E}(T)\mathbb{E}(X_0).$$

As a direct application, we obtain the following.

Lemma 7. $\mathbb{E}(T(M)) = \mathbb{E}(M) \cdot \frac{\mathbb{E}(V_1)}{2}$.

From hitting times to cover time. The following relates the probability to hit a node to the cover time, and can be thought of as a loose but easy Matthew (upper) bound.

Lemma 8. Let $(X_t)_t$ be a finite Markov chain with n states and, for a state x, $M(x)$ be the random number of steps before the chain hits x. If there are m and $p > 0$ such that for every state x, $\Pr(M(x) \leq m) \geq p$, then the cover time of the chain is $O(\frac{m \log(n)}{p})$.

Proof. Split the moves into phases, each composed of m consecutive moves. Using the Markov property of the process, during each phase, x is visited with probability at least p, independently of the trajectory on previous phases. Then the probability to not have visited x after ℓ phases yet, is less than $(1-p)^\ell$. Using a union bound, the probability that there exists a node which has not been visited after ℓ phases is thus less than $n(1-p)^\ell$. For

$$\ell_j = \frac{-\log(2^j n)}{\log(1-p)} = O\left(\frac{j \log n}{p}\right),$$

this probability is less than 2^{-j}. Hence, the expected number of phases before covering all nodes is less than

$$\sum_j \ell_{j+1} \cdot 2^{-j} = O\left(\frac{\log n}{p}\right).$$

Multiplying this number by m gives the expected number of moves before covering all states. □

From the infinite line to the n-nodes cycle.

Remark 9. With the notations of Eq. (1), if for $x \in [0, n-1] \subset \mathbb{Z}$, $M_{\mathbb{Z}}(x)$ (resp. $M_{C_n}(x)$) is the random number of moves for Z (resp $X = Z \mod n$) to hit x, we have $M_{C_n}(x \mod n) \leq M_{\mathbb{Z}}(x)$.

Combining this remark with the two previous lemmas, we obtain:

Claim 10. If we have m and p such that for all $x \in [0, n-1]$, $\Pr(M_{\mathbb{Z}}(x) \leq m) \geq p$, then the cover time of X is $O\left(\frac{\mathbb{E}(V_1) m \log n}{p}\right)$. Note that $\mathbb{E}(V_1) = k$ in the particular base B process with k lengths.

A useful identity. All of our upper bounds rely on the following identity, the usefulness of which was highlighted in [23]. If N is a nonnegative random variable then:

$$\Pr(N \geq 1) = \frac{\mathbb{E}(N)}{\mathbb{E}(N \mid N \geq 1)}. \tag{8}$$

2 Upper Bound on the Cycle

This section is dedicated to proving Theorem 1. We begin with an overview of the proof. Using Claim 10, the bound of Theorem 1 can be established by studying the process on the infinite line.

The core of the computations for Theorem 1 are encapsulated in Lemmas 11 and 12. These bounds on the distribution of the move process Z are used

to lower bound the expected number of visits to any point on the infinite line, and upper bound the expected number of returns to the starting point. A lower bound on the probability of any node being visited follows from Equation (8).

The following two lemmas consider the move process on the infinite line and present the main technical aspects of the proof. The corresponding proofs are all deferred to the full version. Recall, from Theorem 1, that we assume that k is such that $k \geq 2$ and $B^{k-1} \leq n \leq B^k$.

Lemma 11. There exist sufficiently large constants c and c' (independent of all other parameters), such that for any $j \in [0, k-1]$ and $m \geq cB^{j+1} \log kB$, we have $\Pr(Z(m) = 0) \leq c'B^{-j}$.

Lemma 12. Let c (resp. c') be a big (resp. small) enough constant. Let $m_0 := cB^{k+1} \log kB$. For any $x \in [0, n-1] \subseteq [0, B^k - 1]$, and $m \in [m_0, m_0 + n]$, we have $\Pr(Z(m) = x) \geq \frac{c'}{B^k \sqrt{\log kB}}$.

With Claim 10, Lemma 11, and Lemma 12, we are now ready to prove Theorem 1.

Proof (Proof of Theorem 1). Fix a point $x \in \mathbb{Z}$. Define $N_x(m)$ as the number of visits to x after the first m moves have been made. We want to apply Eq. (8) to $N_x(m)$. This will imply a lower bound on the probability that the node x has been visited before m moves (i.e., the event $N_x(m) \geq 1$). Set

$$m_0 = cB^{k+1} \log kB, \quad \text{and} \quad m_1 = m_0 + n,$$

for some constant $c > 0$ big enough so that both the conclusion of Lemma 11 and Lemma 12 hold. Note that since $B^{k-1} \leq n$, $m_1 = O(nB^2 \log kB)$.

Our first goal is to show that $\mathbb{E}(N_x(m_1) \mid N_x(m_1) \geq 1) = O(kB^2 \log kB)$. Since the number of returns to x before time m_1 is maximized when we begin at x, we have:

$$\mathbb{E}(N_x(m_1) \mid N_x(m_1) \geq 1) \leq \mathbb{E}(N_0(m_1)) \leq 1 + \sum_{m=1}^{m_1} \Pr(Z(m) = 0) \ . \tag{9}$$

Next, using Lemma 11, for any $j \in [1, k-1]$ and any m, such that

$$cB^{j+1} \log kB \leq m < cB^{j+2} \log kB,$$

we upper bound $\Pr(Z(m) = 0)$ by $c'B^{-j}$ with c' some other constant. When $m \leq cB^2 \log kB$, we use $\Pr(Z(m) = 0) \leq 1$. Using Inequality (9), we obtain:

$$\begin{aligned} \mathbb{E}(N_x(m_1) \mid N_x(m_1) \geq 1) &= O\left(B^2 \log kB + \sum_{j=1}^{k-2} \frac{B^{j+2} \log kB}{B^j} \right) \\ &= O(kB^2 \log kB) \ , \end{aligned} \tag{10}$$

as desired. We next lower bound $\mathbb{E}\left(N_x(m_1)\right)$ for any $x \in [0, n-1]$, by summing the Inequality of Lemma 12 between m_0 and m_1:

$$\mathbb{E}(N_x(m_1)) \geq \frac{c' \cdot n}{B^k \sqrt{\log kB}} = \Omega\left(\frac{1}{B\sqrt{\log kB}}\right). \tag{11}$$

Dividing (11) by (10), it follows from Eq. (8) that for any $x \in [0, n-1]$, we have $\Pr(N_x(m_1) > 1) = \Omega\left(\frac{1}{kB^3 \log^{3/2} kB}\right)$. Using Claim 10 with $m = m_1 = O(nB^2 \log kB)$ and $p = \Omega\left(\frac{1}{kB^3 \log^{3/2} kB}\right)$, we obtain an upper bound on the expected cover time on the n cycle of $O(k^2 B^5 \log^{5/2} kB \cdot n \log n)$. The result follows by bounding $O(k^2 B^5 \log^{5/2} kB)$ by $\mathrm{poly}(k) \cdot \mathrm{poly}(B)$. □

3 Lower Bound on the Cycle

This section is dedicated to proving Theorem 3. Consider a k-intermittent search X on the cycle C_n and denote by $(p_i)_{i=1}^k$ and $(L_i)_{i=1}^k$ its parameters with $L_i < L_{i+1}$ for all $i \in [k-1]$. We also set $L_{k+1} = n$. Theorem 3 is a direct consequence of the following lemma, as appears in the full version (in short, the biggest multiplicative gap between consecutive L_i is minimized when setting $L_i = n^{\frac{i}{k}}$ and this yields the lower bound).

Lemma 13. There exists a constant $c > 0$ such that, in expectation, for any $i \leq k$, the time needed for X to visit n distinct points is at least

$$c\frac{n}{k} \cdot \sqrt{\frac{L_{i+1}}{L_i}},$$

Proof. Let $i \in [k]$ be fixed throughout the proof. Recall that we need to count time and not the number of moves. We divide time into *phases*, each of length precisely L_{i+1}. We call any jump of length L_j for $j \geq i+1$ a *long jump*. By definition, during a phase, at most one endpoint of a long jump is visited. Let us denote by N_ℓ the number of nodes visited during phase ℓ (some of these nodes may have been previously visited on a phase $\ell' < \ell$). The proof of the following claim appears in the full version.

Claim 14. For every $\ell \in \mathbb{N}$, it holds that $\mathbb{E}(N_\ell) = O\left(k\sqrt{L_i \cdot L_{i+1}}\right)$.

Using Claim 14, we can bound the total number of nodes visited during the first s phases $\mathbb{E}\left(\sum_{\ell=1}^{s} N_\ell\right) \leq s \cdot O\left(k\sqrt{L_i \cdot L_{i+1}}\right)$. Let $s_1 := n \cdot \frac{c}{k \cdot \sqrt{L_i \cdot L_{i+1}}}$ for a small constant c. With this choice for s, the previous bound is less than $n/2$. Using a Markov inequality, we get $\Pr\left(\sum_{\ell=1}^{s_1} N_\ell \geq n\right) < \frac{1}{2}$. Consequently, with probability at least $1/2$, more than s_1 phases are needed in order to visit n distinct nodes. Since each phase lasts (exactly) L_{i+1} time, the total expected time required in order to visit n distinct nodes is at least $s_1 \cdot L_{i+1} = \Omega\left(\frac{n}{k} \cdot \sqrt{\frac{L_{i+1}}{L_i}}\right)$. □

4 Efficient Strategy for Walk or Probe

Our strategy for the *Walk or Probe* problem is simple: instead of probing at every step, we probe every $\sqrt{t_{mix}}$ steps (we omit ceilings for readability). We prove here that this tactic gives the bounds of Theorem 4. In fact, here we essentially prove a bound on the cover time of the Markov chain $(X_{k\sqrt{t_{mix}}})_{k\geq 0}$, where $(X_t)_t$ is the lazy random walk on the connected graph G.

Proof. We first recall some basic results about mixing time. The notion of mixing time we refer to is the total variation mixing time ([29, Section 4.5]). It is defined as:

$$t_{mix} = \min\left\{t \geq 1 : \max_x \sum_y |P^t(x,y) - \pi(y)| \leq \frac{1}{2}\right\},$$

where $P^t(x,\cdot)$ denotes the law of the random walk started at x after t steps and π is the stationary distribution. Lemmas 4.5 and 4.7 in [2] imply that, for $t \geq 4t_{mix}$, $P^t(x,y) \geq \frac{\pi(y)}{2}$. Since $\pi(x) = \Delta(x)/2|E|$, where $\Delta(x)$ is the degree of node x, for $t \geq 4t_{mix}$, we have:

$$\Pr(X_t = x) \geq \frac{\Delta(x)}{4|E|}. \tag{12}$$

Let $\mathcal{N}$ be the number of times we probe x between times $4t_{mix}$ and $(4+c)t_{mix}$. Then, by Eq. (8), the probability that x was probed in this time interval equals

$$\frac{\mathbb{E}(\mathcal{N})}{\mathbb{E}(\mathcal{N}|\mathcal{N} \geq 1)}. \tag{13}$$

By Eq. (12) we have:

$$\mathbb{E}(\mathcal{N}) = \sum_{k=0}^{c\sqrt{t_{mix}}} \Pr(X_{4t_{mix}+k\sqrt{t_{mix}}} = x) \geq c\sqrt{t_{mix}}\frac{\Delta(x)}{4|E|}.$$

For the denominator in (13), since the process is markovian, we can shift the times and so this is equal to $\mathbb{E}\left(N_x(c\,t_{mix})|N_x(c\,t_{mix}) \geq 1)\right)$ where $N_x(c\,t_{mix})$ is the number of times we probe x before time $c\,t_{mix}$. As the number of returns is maximized whenever we begin at x, this is less than $\mathbb{E}\left(N_x(c\,t_{mix})|X_0 = x\right)$. We next use the following bounds on the probability of returns. For any $x \in G$,

$$\Pr(X_t = x) \leq 5/\sqrt{t} \quad \text{if } t \leq 5n^2 \text{ and } G \text{ is regular}, \tag{14}$$

$$\leq \Delta(x)/\sqrt{t} \quad \text{if } t \leq |E|^2 - 1. \tag{15}$$

The bound for regular graphs is taken from Proposition 6.18 in [2], while the general bound follows from the more elaborate bound in Lemma 3.4 in [32]. Let us write these bounds as

$$\Pr(X_t = x) \leq \frac{\beta_x}{\sqrt{t}}$$

with $\beta_x = 5$ if G is regular and $\beta_x = \Delta(x)$ otherwise. Note that we can use these bounds for $t \leq c\, t_{mix}$, for c small enough, since $t_{mix} \leq 3t_{cov}$ ([29, Eq. (10.24)]) and, in connected graphs $t_{cov} \leq 2|E|(n-1) \leq 2|E|^2$ ([3, Theorem]), while in regular graphs we have $t_{cov} \leq 2n^2$ ([17, Corollary 6]). Thus, using that $\sum_{k=1}^{t} \frac{1}{\sqrt{k}} \leq 2\sqrt{t}$, we have:

$$\begin{aligned}\mathbb{E}(N_x(c\, t_{mix})|X_0 = x) &= \sum_{k=0}^{c\sqrt{t_{mix}}} \Pr(X_{k\sqrt{t_{mix}}} = x | X_0 = x) \\ &\leq 1 + \frac{\beta_x}{t_{mix}^{\frac{1}{4}}} \sum_{k=1}^{c\sqrt{t_{mix}}} \frac{1}{\sqrt{k}} \leq 1 + 2\beta_x\sqrt{c}.\end{aligned}$$

Hence, the probability that x is probed before time $(4+c)t_{mix}$ is greater than

$$\frac{\Delta(x)}{4|E|} \frac{c\sqrt{t_{mix}}}{1 + 2\beta_x\sqrt{c}} = \Omega\left(\frac{\sqrt{t_{mix}}}{\gamma}\right),$$

where $\gamma = n$ if G is regular and $\gamma = |E|$ otherwise. By Lemma 8 applied to the Markov chain $(X_{k\sqrt{t_{mix}}})_{k\geq 0}$, the expected number of moves to probe all nodes is $O(\gamma\sqrt{t_{mix}}\log(n))$. Since we probe every $\sqrt{t_{mix}}$ steps with cost C, and a step has a unit cost, the probing cover time is $O(C\gamma\log(n) + \gamma\sqrt{t_{mix}}\log(n))$. This completes the proof of Theorem 4. □

References

1. M. Adler, H. Räcke, N. Sivadasan, C. Sohler, and B. Vöcking. Randomized pursuit-evasion in graphs. *Combinatorics, Probability and Computing*, 12(3):225244, 2003.
2. David Aldous and James Allen Fill. Reversible markov chains and random walks on graphs, 2002. Unfinished monograph, recompiled 2014, available at `http://www.stat.berkeley.edu/aldous/RWG/book.html`.
3. Romas Aleliunas, Richard M. Karp, Richard J. Lipton, Laszlo Lovasz, and Charles Rackoff. Random walks, universal traversal sequences, and the complexity of maze problems. In *SFCS*, Washington, DC, USA, 1979. IEEE Computer Society.
4. F. Bartumeus, J. Catalan, U. L. Fulco, M. L. Lyra, and G. M. Viswanathan. Optimizing the encounter rate in biological interactions: Lévy versus brownian strategies. *Phys. Rev. Lett.*, 88, Feb 2002.
5. Frederic Bartumeus. Lévy processes in animal movement: An evolutionary hypothesis. *Fractals*, 15(02), 2007.
6. Olivier Bénichou, C Loverdo, M Moreau, and R Voituriez. Intermittent search strategies. *Reviews of Modern Physics*, 83(1), 2011.
7. Itai Benjamini, Gady Kozma, and Nicholas Wormald. The mixing time of the giant component of a random graph. *Random Structures and Algorithms*, 45(3):383–407, 2014.
8. Otto G. Berg, Robert B. Winter, and Peter H. Von Hippel. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. models and theory. *Biochemistry*, 20(24), 1981. PMID: 7317363.

9. Denis Boyer et al. Scale-free foraging by primates emerges from their interaction with a complex environment. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1595), 2006.
10. Marie Chupeau, Olivier Benichou, and Raphael Voituriez. Cover times of random searches. *Nature Physics*, 11(10), 2015.
11. Colin Cooper and Alan Frieze. The cover time of the giant component of a random graph. *Random Structures and Algorithms*, 32(4):401–439, 2008.
12. Don Coppersmith, Uriel Feige, and James B. Shearer. Random walks on regular and irregular graphs. *SIAM J. Discrete Math.*, 9(2), 1996.
13. M. Coppey, O. Bnichou, R. Voituriez, and M. Moreau. Kinetics of Target Site Localization of a Protein on DNA: A Stochastic Approach. *Biophysical Journal*, 87(3), 2004.
14. Jurek Czyzowicz, Leszek Gsieniec, Konstantinos Georgiou, Evangelos Kranakis, and Fraser MacQuarrie. The beachcombers' problem: Walking and searching with mobile robots. *Theoretical Computer Science*, 608(Part 3):201 – 218, 2015. Structural Information and Communication Complexity.
15. Andrew M. Edwards et al. Revisiting levy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature*, 449(7165), 2007.
16. A. Einstein. Investigations on the theory of the brownian movement. *Annal. Physik*, 34, 1911.
17. Uriel Feige. Collecting coupons on trees, and the analysis of random walks. Technical report, 1994.
18. Uriel Feige. A tight lower bound on the cover time for random walks on graphs. *Random Structures & Algorithms*, 6(4), 1995.
19. Ehud Fonio, Yael Heyman, Lucas Boczkowski, Aviram Gelblum, Adrian Kosowski, Amos Korman, and Ofer Feinerman. A locally-blazed ant trail achieves efficient collective navigation despite limited information. *eLife*, 5, 2016.
20. Joel Friedman. A proof of alon's second eigenvalue conjecture and related problems. *CoRR*, cs.DM/0405020, 2004.
21. Tajie H Harris et al. Generalized lévy walks and the role of chemokines in migration of effector cd8(+) t cells. volume 486, 2012.
22. Amos Israeli and Marc Jalfon. Token management schemes and random walks yield self-stabilizing mutual exclusion. PODC. ACM, 1990.
23. Varun Kanade, Frederik Mallmann-Trenn, and Thomas Sauerwald. On coalescence time in graphs-when is coalescing as fast as meeting? *CoRR*, abs/1611.02460, 2016.
24. David Kempe, Jon M. Kleinberg, and Alan J. Demers. Spatial gossip and resource location protocols. *J. ACM*, 51(6), 2004.
25. J. Klafter and G. Zumofen. Lévy statistics in a hamiltonian system. *Phys. Rev. E*, 49, Jun 1994.
26. Joseph Klafter, Michael F. Shlesinger, and Gert Zumofen. Beyond brownian motion. In *Physics Today 49, 2, 33*, 1996.
27. Jon M. Kleinberg. The small-world phenomenon: an algorithmic perspective. In *STOC*, 2000.
28. Gregory Lawler and Vlada Limic. *Random Walk: A Modern Introduction*. Cambridge University Press, 2010.
29. David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2008.
30. Michael A. Lomholt et al. Lévy strategies in intermittent search processes are advantageous. *Proceedings of the National Academy of Sciences*, 105(32), 2008.
31. Claude Loverdo. Optimal search strategies and intermittent random walk: from restriction enzymes to the albatross flight. December 2009.

32. Russell Lyons. Asymptotic enumeration of spanning trees. *Combinatorics, Probability and Computing*, 14(4), 2005.
33. G. Oshanin et al. Intermittent random walks for an optimal search strategy: one-dimensional case. *Journal of Physics: Condensed Matter*, 19(6), 2007.
34. K. Pearson. The problem of the random walk. *Nature*, 72(1865), 1905.
35. Omer Reingold. Undirected connectivity in log-space. *J. ACM*, 55(4), 2008.
36. Gretchen Reynolds. Navigating our world like birds and bees. *The New York Times*, 2014.
37. I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong. On the levy-walk nature of human mobility. In *IEEE INFOCOM 2008*.
38. John Travis. Do wandering albatrosses care about math? *Science*, 318(5851), 2007.
39. G. M. Viswanathan et al. Levy flight search patterns of wandering albatrosses. *Nature*, 381(6581), 1996.
40. G. M. Viswanathan et al. Optimizing the success of random searches. *Nature*, 401(6756), 1999.