A Birthday Paradox for Markov chains, with an optimal bound for collision in the Pollard Rho Algorithm for Discrete Logarithm

Jeong Han Kim * Ravi Montenegro † Yuval Peres ‡ Prasad Tetali §

Abstract

We show a Birthday Paradox for self-intersections of Markov chains with uniform stationary distribution. As an application, we analyze Pollard's Rho algorithm for finding the discrete logarithm in a cyclic group G and find that, if the partition in the algorithm is given by a random oracle, then with high probability a collision occurs in $\Theta(\sqrt{|G|})$ steps. This is the first proof of the correct order bound which does not assume that every step of the algorithm produces an i.i.d. sample from G.

1 Introduction

The Birthday Paradox states that if $C\sqrt{N}$ items are sampled uniformly at random, with replacement, from a set of N items, then for large C, with high probability some item will be chosen twice. This can be interpreted as a statement that with high probability, a Markov chain on the complete graph K_N with transitions $P(i,j) = 1/N$ will intersect its past in $C\sqrt{N}$ steps; we refer to such a self-intersection as a *collision*, and say the *"collision time"* is $O(\sqrt{N})$. In [8], this was generalized: for a general Markov chain, the collision time was bounded by $O(\sqrt{N}\,T_s(1/2))$, where $T_s(\epsilon) = \min\{n : \forall u,v \in V,\, P^n(u,v) \geq (1-\epsilon)\pi(v)\}$ measures the time required for the n-step distribution to assign every state a suitable multiple of its stationary probability. In [5], the bound on collision time was improved to $O(\sqrt{N\,T_s(1/2)})$.

The motivation of [8, 5] was to study the collision time for a Markov chain involved in Pollard's Rho algorithm for finding the discrete logarithm on a cyclic group G of prime order $N = |G| \neq 2$. For this walk $T_s(1/2) = \Omega(\log N)$ and so the results of [8, 5] are insufficient to show the widely believed $\Theta(\sqrt{N})$ collision time for this walk. In this paper we improve upon these bounds and show that if a finite ergodic Markov chain has uniform stationary distribution over N states, then $O(\sqrt{N})$ steps suffice for a collision to occur, as long as the relative-pointwise distance (L_∞ of the densities of the current and the stationary distribution) drops steadily *early* in the random walk; it turns out that the precise mixing time is largely, although not entirely, unimportant. See Theorem

*Department of Mathematics, Yonsei University, Seoul, 120-749 Korea, Email: jehkim@yonsei.ac.kr

†Department of Mathematical Sciences, University of Massachusetts at Lowell, Lowell, MA 01854, USA. Email: ravi_montenegro@uml.edu

‡Microsoft Research, Redmond and University of California, Berkeley, CA 94720, USA. Email: peres@microsoft.com; Research supported in part by NSF grant DMS-0605166

§School of Mathematics and School of Computer Science, Georgia Institute of Technology, Atlanta, GA 30332, USA. Email: tetali@math.gatech.edu; research supported in part by NSF grants DMS 0401239, 0701043

3.1 for a precise statement. This is then applied to the Rho walk to give the first proof of collision in $\Theta(\sqrt{N})$ steps.

We note here that it is also well known (see e.g. [1], Section 4.1) that a sample of length L from a Markov chain is roughly equivalent to $L\lambda$ samples from the stationary measure (of the Markov chain) for the purpose of sampling, where λ is the spectral gap of the chain. This yields another estimate on collision time for a Markov chain, which is also of a multiplicative nature (namely, $\sqrt{N}$ *times* a function of the mixing time) as in [8, 5]. A main point of the present work is to establish sufficient criteria under which the collision time has an *additive* bound: $C\sqrt{N}$ plus an estimate on the mixing time. While the Rho algorithm provided the main motivation for the present work, we find the more general Birthday paradox result to be of independent interest, and as such expect to have other applications in the future.

A bit of detail about the Pollard Rho algorithm is in order. The classical discrete logarithm problem on a cyclic group deals with computing the exponents, given the generator of the group; more precisely, given a generator g of a cyclic group G and an element $h = g^x$, one would like to compute x efficiently. Due to its presumed computational difficulty, the problem figures prominently in various cryptosystems, including the Diffie-Hellman key exchange, El Gamal system, and elliptic curve cryptosystems. About 30 years ago, J.M. Pollard suggested algorithms to help solve both factoring large integers [11] and the discrete logarithm problem [12]. While the algorithms are of much interest in computational number theory and cryptography, there has been little work on rigorous analysis. We refer the reader to [8] and other existing literature (e.g., [17, 2]) for further cryptographic and number-theoretical motivation for the discrete logarithm problem.

A standard variant of the classical Pollard Rho algorithm for finding discrete logarithms can be described using a Markov chain on a cyclic group G. While there has been no rigorous proof of rapid mixing of this Markov chain of order $O(\log^c |G|)$ until recently, Miller-Venkatesan [8] gave a proof of mixing of order $O(\log^3 |G|)$ steps and collision time of $O(\sqrt{|G|}\log^3 |G|)$, and Kim et al. [5] showed mixing of order $O(\log |G| \log\log |G|)$ and collision time of $O(\sqrt{|G|\log |G| \log\log |G|})$. In this paper we give the first proof of the correct $\Theta(\sqrt{|G|})$ collision time. By recent results of Miller-Venkatesan [9] this collision will be non-degenerate with probability $1-o(1)$ for almost every prime order $|G|$, if the start point of the algorithm is chosen at random or if there is no collision in the first $O(\log |G| \log\log |G|)$ steps.

The paper proceeds as follows. Section 2 contains some preliminaries; primarily an introduction to the Pollard Rho Algorithm, and a simple multiplicative bound on the collision time in terms of the mixing time. The more general Birthday Paradox for Markov chains with uniform stationary distribution is shown in Section 3. In Section 4 we bound the appropriate constants for the Rho walk and show the optimal collision time. We finish in Section 5 with a few comments on the sharpness of our result.

2 Preliminaries

Our intent in generalizing the Birthday Paradox was to bound the collision time of the Pollard Rho algorithm for Discrete Logarithm. As such, we briefly introduce the algorithm here. Throughout the analysis in the following sections, we assume that the size $N = |G|$ of the cyclic group on which the random walk is performed is odd. Indeed there is a standard reduction – see [13] for a very readable account and also a classical reference [10] – justifying the fact that it suffices to study the discrete logarithm problem on cyclic groups of *prime* order.

Suppose g is a generator of G, that is $G = \{g^i\}_{i=0}^{N-1}$. Given $h \in G$, the discrete logarithm problem asks us to find x such that $g^x = h$. Pollard suggested an algorithm on $\mathbb{Z}_N^\times$ based on a random walk and the Birthday Paradox. A common extension of his idea to groups of prime order is to start with a partition of G into sets S_1, S_2, S_3 of roughly equal sizes, and define an iterating function $F : G \to G$ by $F(y) = gy$ if $y \in S_1$, $F(y) = hy = g^x y$ if $y \in S_2$, and $F(y) = y^2$ if $y \in S_3$. Then consider the walk $y_{i+1} = F(y_i)$. If this walk passes through the same state twice, say $g^{a+xb} = g^{\alpha+x\beta}$, then $g^{a-\alpha} = g^{x(\beta-b)}$ and so $a - \alpha \equiv x(\beta - b) \mod N$ and $x \equiv (a - \alpha)(\beta - b)^{-1} \mod N$, which determines x as long as $(\beta - b, N) = 1$. Hence, if we define a *collision* to be the event that the walk passes over the same group element twice, then the first time there is a collision it might be possible to determine the discrete logarithm.

To estimate the running time until a collision, one heuristic is to treat F as if it outputs uniformly random group elements. By the Birthday Paradox if $O(\sqrt{|G|})$ group elements are chosen uniformly at random, then there is a high probability that two of these are the same. Teske [16] has given experimental evidence that the time until a collision is slower than what would be expected by an independent uniform random process. We analyze instead the actual Markov chain in which it is assumed only that each $y \in G$ is assigned independently and at random to a partition S_1, S_2 or S_3. In this case, although the iterating function F described earlier is deterministic, because the partition of G was randomly chosen then the walk is equivalent to a Markov chain (i.e. a random walk), at least until the walk visits a previously visited state and a collision occurs. The problem is then one of considering a walk on the exponent of g, that is a walk P on the cycle $\mathbb{Z}_N$ with transitions $P(u, u + 1) = P(u, u + x) = P(u, 2u) = 1/3$.

Remark 2.1. By assuming each $y \in G$ is assigned independently and at random to a partition we have eliminated one of the key features of the Pollard Rho algorithm, space efficiency. However, if the partitions are given by a hash function $f : (G, N) \to \{1, 2, 3\}$ which is sufficiently pseudo-random then we might expect behavior similar to the model with random partitions.

Remark 2.2. While we are studying the time until a collision occurs, there is no guarantee that the first collision will be non-degenerate. If the first collision is degenerate then so also will be all collisions, as the algorithm becomes deterministic after the first collision.

As mentioned in the introduction, we first recall a simple multiplicative bound on collision time from [5]. The following proposition relates $T_s(1/2)$ to the time until a collision occurs for any Markov chain P with uniform distribution on G as the stationary distribution.

Proposition 2.3. *With the above definitions, a collision occurs after*

$$1 + T_s(1/2) + 2\sqrt{2c\,|G|\,T_s(1/2)}$$

steps, with probability at least $1 - e^{-c}$*, for any* $c > 0$.

Proof. Let S denote the first $\left\lceil \sqrt{2c\,|G|\,T_s(1/2)} \right\rceil$ states visited by the walk. If two of these states are the same then a collision has occurred, so assume all states are distinct. Even if we only check for collisions every $T_s(1/2)$ steps, the chance that no collision occurs in the next $tT_s(1/2)$ steps (so consider t semi-random states) is then at most

$$\left(1 - \frac{1}{2}\frac{|S|}{|G|}\right)^t \le \left(1 - \sqrt{\frac{c\,T_s(1/2)}{2|G|}}\right)^t \le \exp\left(-t\sqrt{\frac{c\,T_s(1/2)}{2|G|}}\right).$$

When $t = \left\lceil \sqrt{\frac{2c|G|}{T_s(1/2)}} \right\rceil$, this is at most e^{-c}, as desired, and so at most

$$\left\lceil \sqrt{2c\,|G|\,T_s(1/2)} \right\rceil + \left\lceil \sqrt{\frac{2c|G|}{T_s(1/2)}} \right\rceil \, T_s(1/2)$$

steps are required for a collision to occur with probability at least $1 - e^{-c}$. □

Obtaining a more refined additive bound on collision time will be the focus of the next section. While the proof can be seen as another application of the well-known second moment method, it turns out that bounding the second moment of the number of collisions *before* the mixing time is somewhat subtle. To handle this, we use an idea from [6], who in turn credit their line of calculation to [4].

3 Collision Time

Consider a finite ergodic Markov chain P with uniform stationary distribution (i.e. doubly stochastic), state space Ω of cardinality $N = |\Omega|$, and let $X_0, X_1, \cdots$ denote a particular instance of the walk. In this section we determine the number of steps of the walk required to have a high probability that a "collision" has occurred, i.e. a self-intersection $X_i = X_j$ for some $i \neq j$.

First, some notation. Fix some $T \geq 0$. Define

$$S = \sum_{i=0}^{\beta\sqrt{N}} \sum_{j=i+2T}^{\beta\sqrt{N}+2T} 1_{\{X_i = X_j\}}$$

to be the number of times the walk intersects itself in $\beta\sqrt{N} + 2T$ steps, where i and j are at least $2T$ steps apart. Also, for $u, v \in \Omega$, let

$$G_T(u,v) = \sum_{i=0}^{T} P^i(u,v)$$

be the expected number of times a walk beginning at u hits state v in T steps. Finally, let

$$A_T = \max_u \sum_v G_T^2(u,v) \qquad \text{and} \qquad A_T^* = \max_u \sum_v G_T^2(v,u)\,.$$

To see the connection between these and the collision time, observe that

$$\begin{aligned}
\sum_v G_T^2(u,v) &= \sum_v \Big(\sum_{i=0}^{T} \sum_{j=0}^{T} P^i(u,v) P^j(u,v) \Big) \\
&= \sum_{i=0}^{T} \sum_{j=0}^{T} \sum_v P^i(u,v) P^j(u,v) \\
&= \sum_{i=0}^{T} \sum_{j=0}^{T} \mathbb{P}_{u,u}(X_i = Y_j) \\
&= \sum_{i=0}^{T} \sum_{j=0}^{T} E\big(\mathbf{1}_{\{X_i = Y_j\}}\big) = E \sum_{i,j=0}^{T} \mathbf{1}_{\{X_i = Y_j\}}\,,
\end{aligned}$$

where $\{X_i\}, \{Y_j\}$ are i.i.d. copies of the chain, both having started at u at time 0. Hence A_T is the maximal expected number of collisions of two T-step i.i.d. walks of P starting at the same state u, while A_T^* is the same for P^*.

The main result of this section is the following.

Theorem 3.1 (Birthday Paradox for Markov chains). *Consider a finite ergodic Markov chain with uniform stationary distribution on a state space of N vertices. Let T be such that $\frac{m}{N} \leq P^T(u,v) \leq \frac{M}{N}$ for some $m \leq 1 \leq M$ and every pair of states u, v. After*

$$4c\left(\frac{M}{m}\right)^2\left(\sqrt{\frac{2N}{M}\max\{A_T,\ A_T^*\}} + T\right)$$

steps a collision occurs with probability at least $1 - e^{-c}$, for any $c \geq 0$.

Proof. First recall the standard second moment bound: using Cauchy-Schwarz, we have that

$$E[S] = E[S\mathbf{1}_{\{S>0\}}] \leq E[S^2]^{1/2}E[\mathbf{1}_{\{S>0\}}]^{1/2}$$

and hence $\Pr[S > 0] \geq E[S]^2/E[S^2]$. By Lemma 3.3, if $\beta = 2\sqrt{2\max\{A_T, A_T^*\}/M}$ then

$$\Pr[S > 0] \geq \frac{m^2/M^2}{1 + \frac{8\max\{A_T,A_T^*\}}{M\beta^2}} \geq \frac{m^2}{2M^2}\,, \tag{3.1}$$

independent of the starting point. Hence the probability that there is no collision after $k(\beta\sqrt{N}+2T)$ steps is at most $(1 - m^2/2M^2)^k \leq e^{-km^2/2M^2}$. Taking $k = 2cM^2/m^2$ completes the proof. □

Remark 3.2. Observe that if $A_T, A_T^*, m, M = \Theta(1)$ and $T = O(\sqrt{N})$ then the collision time is $O(\sqrt{N})$, as in the standard Birthday Paradox. By Lemma 3.4, it will suffice that P^T be sufficiently close to uniform after $T = o(\sqrt{N})$ steps, and that $P^j(u,v) = o(T^{-2}) + d^j$ for all u, v, for $j \leq T$ and some $d < 1$.

When applied to the standard Birthday Paradox equation (3.1) with $T = 1$ is $2/\sqrt{\ln 2} \approx 2.4$ times the correct number of steps required to reach probability 1/2. In the final section of the paper, we present an example to illustrate the need for the pre-mixing term A_T in Theorem 3.1. A slight strengthening of Theorem 3.1 is also shown there, at the cost of a somewhat less intuitive bound.

The proof of Theorem 3.1 relied largely on the following:

Lemma 3.3. *Under the conditions of Theorem 3.1,*

$$E[S] \geq \frac{m}{N}\binom{\beta\sqrt{N}+2}{2}, \qquad E[S^2] \leq \frac{M^2}{N^2}\binom{\beta\sqrt{N}+2}{2}^2\left(1 + \frac{8\max\{A_T, A_T^*\}}{M\beta^2}\right).$$

Proof. We will repeatedly use the relation that there are $\binom{\beta\sqrt{N}+2}{2}$ choices for i, j appearing in the summation for S, i.e. $0 \leq i$ and $i + 2T \leq j \leq \beta\sqrt{N} + 2T$.

Now to the proof. The expectation $E[S]$ satisfies

$$E[S] = E\sum_{i=0}^{\beta\sqrt{N}}\sum_{j=i+2T}^{\beta\sqrt{N}+2T} 1_{\{X_i=X_j\}} = \sum_{i=0}^{\beta\sqrt{N}}\sum_{j=i+2T}^{\beta\sqrt{N}+2T} E[1_{\{X_i=X_j\}}] \geq \binom{\beta\sqrt{N}+2}{2}\frac{m}{N}$$

because if $j \geq i+T$ then

$$Pr(X_j = X_i) = \sum_u Pr(X_i = u)P^{j-i}(u,u) \geq \sum_u Pr(X_i = u)\frac{m}{N} = \frac{m}{N}\,.$$

Similarly, $Pr(X_j = X_i) \leq \frac{M}{N}$ when $j \geq i+T$.

Now for $E[S^2]$. Note that

$$\begin{aligned} E[S^2] &= E\left(\sum_{i=0}^{\beta\sqrt{N}}\sum_{j=i+2T}^{\beta\sqrt{N}+2T} 1_{\{X_i=X_j\}}\right)\left(\sum_{k=0}^{\beta\sqrt{N}}\sum_{l=k+2T}^{\beta\sqrt{N}+2T} 1_{\{X_k=X_l\}}\right) \\ &= \sum_{i=0}^{\beta\sqrt{N}}\sum_{k=0}^{\beta\sqrt{N}}\sum_{j=i+2T}^{\beta\sqrt{N}+2T}\sum_{l=k+2T}^{\beta\sqrt{N}+2T} Prob(X_i = X_j,\, X_k = X_l)\,. \end{aligned}$$

To evaluate this quadruple sum we break it into 3 cases.

Case 1: Suppose $|j-l| \geq T$. Without loss, assume $l \geq j$, so in particular $l \geq \max\{i,j,k\}+T$. Then

$$\begin{aligned} Prob(X_i = X_j,\, X_k = X_l) &= Prob(X_i = X_j)\, Prob(X_l = X_k \mid X_i = X_j) \\ &\leq Prob(X_i = X_j)\, \max_{u,v} Prob(X_l = v \mid X_{\max\{i,j,k\}} = u) \\ &\leq Prob(X_i = X_j)\,\frac{M}{N} \leq \left(\frac{M}{N}\right)^2 . \end{aligned}$$

The first inequality is because $\{X_t\}$ is a Markov chain and so given X_i, X_j, X_k the walk at any time $t \geq \max\{i,j,k\}$ depends only on the state $X_{\max\{i,j,k\}}$.

Case 2: Suppose $|i-k| \geq T$ and $|j-l| < T$. Without loss, assume $i \leq k$. If $j \leq l$ then

$$\begin{aligned} Prob(X_i = X_j,\, X_k = X_l) &= \sum_{u,v} Prob(X_i = u)\, P^{k-i}(u,v)P^{j-k}(v,u)P^{l-j}(u,v) \\ &\leq \sum_u Prob(X_i = u)\,\frac{M}{N}\,\frac{M}{N}\sum_v P^{l-j}(u,v) = \left(\frac{M}{N}\right)^2 \end{aligned}$$

because $k \geq i+T$, $j \geq k+T$, and $\sum_v P^t(u,v) = 1$ for any t because P and hence also P^t is a stochastic matrix. If, instead, $l < j$ then essentially the same argument works, but with $\sum_v P^t(v,u) = 1$ because P and hence also P^t is doubly-stochastic.

Case 3: Finally, consider those terms with $|j-l| < T$ and $|i-k| < T$. Without loss, assume $i \leq k$. If $l \leq j$ then

$$\begin{aligned} Prob(X_i = X_j,\, X_k = X_l) &= \sum_{u,v} Prob(X_i = u) P^{k-i}(u,v) P^{l-k}(v,v) P^{j-l}(v,u) \\ &\leq \sum_u Prob(X_i = u) \sum_v P^{k-i}(u,v) \frac{M}{N} P^{j-l}(v,u)\,. \end{aligned}$$

The sum over elements with $i \leq k < i+T$ and $l \leq j < l+T$ is upper bounded as follows:

$$\begin{aligned} &\sum_{i=0}^{\beta\sqrt{N}} \sum_{k=i}^{i+T} \sum_{l=k+2T}^{\beta\sqrt{N}+2T} \sum_{j=l}^{l+T} Prob(X_i = X_j,\, X_k = X_l) \\ &\leq \frac{M}{N} \sum_{i=0}^{\beta\sqrt{N}} \sum_{l=i+2T}^{\beta\sqrt{N}+2T} \max_u \sum_v \sum_{k\in[i,i+T)} P^{k-i}(u,v) \sum_{j\in[l,l+T)} P^{j-l}(v,u) \qquad (3.2) \\ &\leq \frac{M}{N} \sum_{i=0}^{\beta\sqrt{N}} \sum_{l=i+2T}^{\beta\sqrt{N}+2T} \max_u \sum_v G_T(u,v) G_T(v,u) \\ &\leq \frac{M}{N} \sum_{i=0}^{\beta\sqrt{N}} \sum_{l=i+2T}^{\beta\sqrt{N}+2T} \max_u \sqrt{\sum_v G_T^2(u,v) \sum_v G_T^2(v,u)} \\ &\leq \frac{M}{N} \binom{\beta\sqrt{N}+2}{2} \sqrt{A_T\, A_T^*}\,. \end{aligned}$$

The case when $j < l$ gives the same bound, but with the observation that $j \geq k+T$ and with A_T instead of $\sqrt{A_T\, A_T^*}$.

Putting together these various cases we get that

$$E[S^2] \leq \binom{\beta\sqrt{N}+2}{2}^2 \left(\frac{M}{N}\right)^2 + 2\binom{\beta\sqrt{N}+2}{2} \frac{M}{N} A_T + 2\binom{\beta\sqrt{N}+2}{2} \frac{M}{N} \sqrt{A_T\, A_T^*}\,.$$

The $\binom{\beta\sqrt{N}+2}{2}^2$ term is the total number of values of i,j,k,l appearing in the sum for $E[S^2]$, and hence also an upper bound on the number of values in Cases 1 and 2. Along with the relation $\binom{\beta\sqrt{N}+2}{2} \geq \frac{\beta^2 N}{2}$ this simplifies to complete the proof. □

To upper bound A_T and A_T^* it suffices to show that the maximum probability of being at a vertex decreases quickly.

Lemma 3.4. *If a finite ergodic Markov chain has uniform stationary distribution then*

$$A_T, A_T^* \leq 2\sum_{j=0}^{T} (j+1) \max_{u,v} P^j(u,v)\,.$$

Proof. If u is such that equality occurs in the definition of A_T, then

$$\begin{aligned} A_T &= \sum_v G_T^2(u,v) = \sum_{i=0}^{T}\sum_{j=0}^{T}\sum_v P^i(u,v)P^j(u,v) \\ &\le 2\sum_{j=0}^{T}\sum_{i=0}^{j}\max_y P^j(u,y)\sum_v P^i(u,v) \\ &\le 2\sum_{j=0}^{T}(j+1)\max_y P^j(u,y)\,. \end{aligned}$$

The same bound holds for A_T^*, which plays the role of A_T for the reversed chain, because the upper bound just shown is the same for the chain and its reversal. □

In particular, suppose $P^j(u,v) \le c + d^j$ for every $u,v \in \Omega$ and some $c,d \in [0,1)$. The sum

$$\begin{aligned} \sum_{j=0}^{T}(j+1)(c+d^j) &= c\,\frac{(T+1)(T+2)}{2} + \frac{1-d^{T+1}-(T+1)d^{T+1}(1-d)}{(1-d)^2} \\ &\le (1+o(1))\frac{cT^2}{2} + \frac{1}{(1-d)^2}\,, \end{aligned}$$

and so if $P^j(u,v) \le o(T^{-2}) + d^j$ for every $u,v \in \Omega$ then $A_T, A_T^* = \frac{2+o(1)}{(1-d)^2}$.

4 Convergence of the Rho walk

Let us now turn our attention to the Pollard Rho walk for discrete logarithm. To apply the collision time result we will first show that $\max_{u,v\in\mathbb{Z}_N} P^s(u,v)$ decreases quickly in s so that Lemma 3.4 may be used. We then find T such that $P^T(u,v) \approx 1/N$ for every $u,v \in \mathbb{Z}_N$. However, instead of studying the Rho walk directly, most of the work will instead involve a "block walk" in which only a certain subset of the states visited by the Rho walk are considered.

Definition 4.1. Let us refer to the three types of moves that the Pollard Rho random walk makes, namely $(u,u+1),(u,u+x)$, and $(u,2u)$, as moves of Type 1, Type 2, and Type 3, respectively. In general, let the random walk be denoted by $Y_0, Y_1, Y_2, \ldots$, with Y_t indicating the position of the walk (modulo N) at time $t \ge 0$. Let T_1 be the first time that the walk makes a move of Type 3. Let $b_1 = Y_{T_1-1} - Y_{T_0}$ (i.e., the ground covered, modulo N, only using consecutive moves of Types 1 and 2.) More generally, let T_i be the first time, since T_{i-1}, that a move of Type 3 happens and set $b_i = Y_{T_i-1} - Y_{T_{i-1}}$. Then the *block walk* B is the walk $X_s = Y_{T_s} = 2^s Y_{T_0} + 2\sum_{i=1}^{s} 2^{s-i} b_i$. Also, for $\delta \in [0,1]$ the $(1+\delta)$-block walk has transition matrix $\mathsf{B}_{1+\delta} = (1-\delta)\mathsf{B} + \delta\mathsf{B}^2$.

By combining our Birthday Paradox for Markov chains with several lemmas to be shown in this section we obtain the main result of the paper:

Theorem 4.2. *For every choice of starting state, the expected number of steps required for the Pollard Rho algorithm for discrete logarithm on a group G to have a collision is at most*

$$(1+o(1))\,12\sqrt{19}\,\sqrt{|G|} < (1+o(1))\,52.5\,\sqrt{|G|}\,.$$

Proof. We work with Theorem 5.2, shown in the Concluding Remarks, because this gives a somewhat sharper bound. Alternatively, Theorem 3.1 and Lemma 3.4 can be applied nearly identically to get the slightly weaker $(1+o(1))72\sqrt{|G|}$.

First consider steps of the $(1+\delta)$-block walk with $\delta = 1/\log_2 N$. Note that $\mathsf{B}^s_{1+\delta}(u,v) \leq \max_{k\in[s,2s]} \mathsf{B}^k(u,v)$, and so Lemma 4.3 implies that $\mathsf{B}^s_{1+\delta}(u,v) \leq \frac{3/2}{\sqrt{N}} + \left(\frac{2}{3}\right)^s$, for $s \geq 0$, and for all u,v. Hence, by equation (5.4), if $T = o(\sqrt[4]{N})$ then $1 + \sum_{j=1}^{2T} 3j\, P^j(u,v) \leq 19 + o(1)$. By Lemma 4.5, after $T = 500(\log_2^4 N) = o(\sqrt[4]{N})$ steps, we have $M \leq 1 + 1/N^2$ and $m \geq 1 - 1/N^2$. Plugging this into Theorem 5.2, a collision fails to occur in

$$k\left(2\sqrt{\left(1+\sum_{j=1}^{2T} 3j \max_{u,v} P^j(u,v)\right)\frac{N}{M}} + 2T\right) = (1+o(1))\, 2\sqrt{19}\, k\sqrt{N}$$

steps with probability at most $(1-\delta)^k$ where $\delta = m^2/2M^2 = (1-o(1))/2$. By Chebyshev's Inequality this requires $(1+o(1))^2\, 2\sqrt{19}\, k\sqrt{N}$ steps of the Block walk with probability $1-o(1)$, and so in $(1+o(1))\, 2\sqrt{19}\, k\sqrt{N}$ steps of the Block walk there is a collision with probability $\frac{1-o(1)}{2}$.

Now let us return to the Rho walk. Recall that T_i denotes the number of Rho steps required for i block steps. The difference $T_{i+1} - T_i$ is an i.i.d. random variable with the same distribution as $T_1 - T_0$. Hence, if $i \geq j$ then $E[T_i - T_j] = (i-j)\, E[T_1 - T_0] = 3(i-j)$. In particular, if we let $r = (1+o(1))\, 2\sqrt{19\, N}$, let R denote the number of Rho steps before a collision, and let B denote the number of block steps before a collision, then

$$\begin{aligned} E[R] &\leq \sum_{k=0}^{\infty} Pr[B > kr]\, E[T_{(k+1)r} - T_{kr} \mid B > kr] \\ &= \sum_{k=0}^{\infty} Pr[B > kr]\, E[T_{(k+1)r} - T_{kr}] \\ &\leq \sum_{k=0}^{\infty} \left(\frac{1+o(1)}{2}\right)^k 3r = (1+o(1))\, 12\sqrt{19}\,\sqrt{N}\,. \end{aligned}$$

□

Now to the first lemma required for the collision bound, a proof that $\mathsf{B}^s(u,v)$ decreases quickly for the block walk:

Lemma 4.3. *If $s \leq \lfloor \log_2 N \rfloor$ then for every $u,v \in \mathbb{Z}_N$ the block walk satisfies*

$$\mathsf{B}^s(u,v) \leq (2/3)^s\,.$$

If $s > \lfloor \log_2 N \rfloor$ then $\mathsf{B}^s(u,v) \leq \dfrac{3/2}{N^{1-\log_2 3}} \leq \dfrac{3/2}{\sqrt{N}}$.

Proof. We start with a weaker, but somewhat more intuitive, proof of a bound on $\mathsf{B}^s(u,v)$ and then improve it to obtain the result of the lemma. The key idea here will be to separate out a portion of the Markov chain which is tree-like with some large depth L, namely the moves induced solely by $b_i = 0$ and $b_i = 1$ moves. Because of the high depth of the tree, the walk spreads out for the first L steps, and hence the probability of being at a vertex also decreases quickly.

Let $S = \{i \in [1 \ldots s] : b_i \in \{0,1\}\}$ and $z = \sum_{i \notin S} 2^{s-i} b_i$. Then $Y_{T_s} = 2^s Y_{T_0} + 2z + 2\sum_{i\in S} 2^{s-i} b_i$. Hence, choosing $Y_{T_0} = u$, $Y_{T_s} = v$, we may write

$$\begin{aligned}
\mathsf{B}^s(u,v) &= \sum_S Prob(S) \sum_{z\in\mathbb{Z}_N} Prob(z \mid S)\, Prob\left(\sum_{i\in S} 2^{s-i} b_i = v/2 - 2^{s-1}u - z \mid z, S\right) \\
&\le \sum_S Prob(S) \max_{w\in\mathbb{Z}_N} Prob\left(\sum_{i\in S} 2^{s-i} b_i = w \mid S\right),
\end{aligned}$$

and so for a fixed choice of S, we can ignore what happens on S^c.

Each $w \in [0 \ldots N-1]$ has a unique binary expansion, and so if $s \le \lfloor \log_2 N \rfloor$ then modulo N each w can still be written in at most one way as an s bit string. For the block walk, $Prob(b_i = 0) \ge 1/3$ and $Prob(b_i = 1) \ge 1/9$, and so $\max\{Prob(b_i = 0 \mid i \in S),\, Prob(b_i = 1 \mid i \in S)\} \le \frac{8}{9}$. It follows that

$$\max_{w\in\mathbb{Z}_N} Prob\left(\sum_{i\in S} 2^{s-i} b_i = w \mid S\right) \le (8/9)^{|S|}, \tag{4.3}$$

using independence of the b_i's. Hence,

$$\begin{aligned}
\mathsf{B}^s(u,v) &\le \sum_S Prob(S)\,(8/9)^{|S|} = \sum_{r=0}^{s} Prob(|S| = r)\,(8/9)^r \\
&\le \sum_{r=0}^{s} \binom{s}{r} \left(\frac{4}{9}\right)^r \left(1 - \frac{4}{9}\right)^{s-r} \left(\frac{8}{9}\right)^r = \left(\frac{4}{9}\frac{8}{9} + \frac{5}{9}\right)^s = \left(\frac{77}{81}\right)^s.
\end{aligned}$$

The second inequality was because $(8/9)^{|S|}$ is decreasing in $|S|$ and so underestimating $|S|$ by assuming $Prob(i \in S) = 4/9$ will only increase the upper bound on $\mathsf{B}^s(u,v)$.

In order to improve on this, we will shortly re-define S (namely, events $\{i \in S\}, \{i \notin S\}$) and auxiliary variables c_i, using the steps of the Rho walk. Also note that the block walk is induced by a Rho walk, so we may assume that the b_i were constructed by a series of steps of the Rho walk. With probability $1/4$ set $i \in S$ and $c_i = 0$, otherwise if the first step is of Type 1 then set $i \in S$ and $c_i = 1$, while if the first step is of Type 3 then put $i \notin S$ and $c_i = 0$, and finally if the first step is of Type 2, then again repeat the above decision making process, using the subsequent steps of the walk. Note that the above construction can be summarized as consisting of one of four equally likely outcomes (at each time), where the last three outcomes depend on the type of the step that the Rho walk takes; indeed each of these three outcomes happens with probability $\frac{3}{4} \times \frac{1}{3} = 1/4$; finally, a Type 2 step forces us to reiterate the four-way decision making process.

Then $Pr(i \in S) = \sum_{l=0}^{\infty} (1/4)^l\,(1/2) = 2/3$. Also observe that $Pr(c_i = 0 | i \in S) = Pr(c_i = 1 | i \in S)$, and that $Pr(b_i - c_i = x \mid i \in S,\, c_i = 0) = Pr(b_i - c_i = x \mid i \in S,\, c_i = 1)$. Hence the steps done earlier (leading to the weaker bound) carry through with $z = \sum_i 2^{s-i}(b_i - c_i)$ and with $\sum_{i\in S} 2^{s-i} b_i$ replaced by $\sum_{i\in S} 2^{s-i} c_i$. In (4.3) replace $(8/9)^{|S|}$ by $(1/2)^{|S|}$, and in showing the final upper bound on $\mathsf{B}^s(u,v)$ replace $4/9$ by $2/3$. This leads to the bound $\mathsf{B}^s(u,v) \le (2/3)^s$.

Finally, when $s > \lfloor \log_2 N \rfloor$, simply apply the preceding argument to $S' = S \cap [1 \ldots \lfloor \log_2 N \rfloor]$. Alternately, note that when $s \ge \lfloor \log_2 N \rfloor$ then $\mathsf{B}^s(u,v) \le \max_w \mathsf{B}^{\lfloor \log_2 N \rfloor}(u,w)$, for every doubly-stochastic Markov chain B. □

In order to use the Birthday Paradox on the Rho walk it suffices to show a mixing time bound of $T = O(\sqrt[4]{N})$ (to guarantee that $A_T, A_T^* = O(1)$). The first such bound was shown by Miller and Venkatesan [8] using characters and quadratic forms, albeit for the Rho walk rather than the Block walk; other sufficiently strong bounds are shown in [5] using canonical paths or Fourier analysis. The argument given here is chosen for brevity alone.

Perhaps the most widely used approach to bounding mixing times is the method of canonical paths. Canonical path methods [15] can be used to lower bound the spectral gap of a Markov kernel P in terms of paths involving edges of P. Fill [3], building on work of Mihail [7], showed a bound on the mixing time in terms of the smallest singular value of P, or equivalently the spectral gap of $\mathsf{P}\mathsf{P}^*$, where the time-reversed walk is $\mathsf{P}^*(v,u) = \frac{\pi(u)\mathsf{P}(u,v)}{\pi(v)} = \mathsf{P}(u,v)$, when the stationary distribution π is uniform. By combining these two methods we obtain a bound on mixing time in terms of even length paths alternating between edges of P and P^*.

Theorem 4.4. *Consider a finite Markov chain P on state space V with stationary distribution π, and set $\pi_* = \min_{v\in V} \pi(v)$. For every $u, v \in V$, $u \neq v$, define a path γ_{uv} from u to v along edges of $\mathsf{P}\mathsf{P}^*$, and let*

$$A = A(\Gamma) = \max_{x\neq y:\mathsf{P}(x,y)\neq 0} \frac{1}{\pi(x)\mathsf{P}(x,y)} \sum_{a\neq b:(x,y)\in\gamma_{ab}} \pi(a)\pi(b)|\gamma_{ab}|\,.$$

Then, for every $u, v \in V$,

$$T \geq 2\,A\,\log\frac{1}{\epsilon\pi_*} \qquad \Longrightarrow \qquad \left|\frac{\mathsf{P}^T(u,v)}{\pi(v)} - 1\right| \leq \epsilon\,.$$

To apply this we need only construct an appropriate set of paths for the $(1+\delta)$-block walk:

Lemma 4.5. *If $T \geq \frac{486}{\delta(1-\delta)}\lceil\log_2 N\rceil^3$ then $\left|\frac{\mathsf{B}^T_{1+\delta}(u,v)}{\pi(v)} - 1\right| \leq \frac{1}{N^2}$ for every $u, v \in \mathbb{Z}_N$.*

Proof. We will construct paths and apply Theorem 4.4 with $\epsilon = \pi_* = 1/N$. If $u \in \mathbb{Z}_N$ then

$$\mathsf{B}_{1+\delta}\mathsf{B}^*_{1+\delta}(u, 2u+1) \geq \mathsf{B}_{1+\delta}(u, 4u+2)\mathsf{B}^*_{1+\delta}(4u+2, 2u+1) \geq \frac{\delta}{27}\,\frac{1-\delta}{3} = \frac{\delta(1-\delta)}{81}\,,$$

and likewise $\mathsf{B}_{1+\delta}\mathsf{B}^*_{1+\delta}(u, 2u) \geq \mathsf{B}_{1+\delta}(u, 4u)\mathsf{B}^*_{1+\delta}(4u, 2u) \geq \frac{\delta}{9}\frac{1-\delta}{3} \geq \frac{\delta(1-\delta)}{81}$.

To construct a path from u to v, set $n = \lceil\log_2 N\rceil$ and $x = (v - 2^n u) \mod N$. Then x has a unique n-bit binary expansion $x = x_0x_1\cdots x_{n-2}x_{n-1}$. To describe the path let $u_0 = u$ and inductively define $u_{i+1} = 2u_i + x_i$. Then $u_n \equiv 2^n u + x \equiv v \mod N$ and $|\gamma_{uv}| = n$.

It remains to count the number of paths through each edge. Fix edge (a,b) with $b \equiv 2a \mod N$ or $b \equiv 2a+1 \mod N$. There are 2^{i-1} potential values of u, and 2^{n-i} potential values of v, such that (a,b) is the i-th edge of path γ_{uv}, and there are n potential values for i, for a total of at most $n\,2^{n-1} \leq n\,N$ paths passing through edge (a,b). □

5 Concluding Remarks

As promised in Section 3, we now present an example that illustrates the need for the pre-mixing term A_T in Theorem 3.1.

Example 5.1. Consider the random walk on $\mathbb{Z}_N$ which transitions from $u \to u+1$ with probability $1 - 1/\sqrt{N}$, and with probability $1/\sqrt{N}$ transitions $u \to v$ for a uniformly random choice of v.

Heuristically the walk proceeds as $u \to u+1$ for $\approx \sqrt{N}$ steps, then randomizes, then proceeds as $u \to u+1$ for another $\sqrt{N}$ steps. This effectively splits the state space into $\sqrt{N}$ blocks of size about $\sqrt{N}$ each, so by the standard Birthday Paradox it should require about $\sqrt{N^{1/2}}$ of these randomizations before a collision will occur. In short, about $N^{3/4}$ steps in total.

To see the need for the pre-mixing term, observe that $T_s \approx \sqrt{N}\log 2$ while if $T = T_\infty \approx \sqrt{N}\log(2(N-1))$ then we may take $m = 1/2$ and $M = 3/2$ in Theorem 3.1. So, whether T_s or T_∞ are considered, it will be insufficient to take $O(T + \sqrt{N})$ steps. However, the number A_T of collisions between two independent copies of this walk is about $\sqrt{N}$, since once a randomization step occurs then the two independent walks are unlikely to collide anytime soon. Our collision time bound says that $O(N^{3/4})$ steps will suffice, which is the correct bound.

A proper analysis shows that $\frac{1-o(1)}{\sqrt{2}}N^{3/4}$ steps are necessary to have a collision with probability 1/2. Conversely, when $T = \sqrt{N}\log^2 N$ then $m = 1 - o(1)$, $M = 1 + o(1)$ and $A_T, A_T^* \leq \frac{1+o(1)}{2}\sqrt{N}$, so by equation (3.1), $(2 + o(1))N^{3/4}$ steps are sufficient to have a collision with probability at least 1/2. Our upper bound is thus off by at most a factor of $2\sqrt{2} \approx 2.8$.

Also, the slight sharpening that was used to derive our improved bound for the Pollard Rho walk:

Theorem 5.2 (Improved Birthday paradox). *Consider a finite ergodic Markov chain with uniform stationary distribution on a state space of N vertices. Let T be such that $\frac{m}{N} \leq P^T(u,v) \leq \frac{M}{N}$ for some $m \leq 1 \leq M$ and every pair of states u, v. After*

$$2c\left(\sqrt{\left(1 + \sum_{j=1}^{2T} 3j \max_{u,v} P^j(u,v)\right)\frac{N}{M}} + T\right)$$

steps a collision occurs with probability at least $1 - \left(1 - \frac{m^2}{2M^2}\right)^c$, independent of the starting state.

Proof. We give only the steps that differ from before. First, in equation (3.2), note that the triple sum after $\max_u$ can be re-written as

$$\sum_{\alpha\in[0,T)}\sum_{\beta\in[0,T)}\sum_v P^\alpha(u,v)P^\beta(v,u) \leq \sum_{\gamma=0}^{2(T-1)} (\gamma+1)P^\gamma(u,u)$$

and so the original quadruple sum reduces to $\frac{M}{N}\binom{\beta\sqrt{N}+2}{2}\max_u \sum_{\gamma=0}^{2(T-1)}(\gamma+1)P^\gamma(u,u)$.

For the case when $i < k$ and $j < l$ proceed similarly, then reduce as in Lemma 3.4 to obtain the upper bound

$$\frac{M}{N}\binom{\beta\sqrt{N}+2}{2}\sum_{\alpha=1}^{T-1}\sum_{\beta=1}^{T-1}\sum_v P^\alpha(u,v)P^\beta(u,v) \leq \frac{M}{N}\binom{\beta\sqrt{N}+2}{2}\sum_{\gamma=1}^{T-1}(2\gamma-1)\max_v P^\gamma(u,v)\,.$$

Adding these two expressions gives an expression of at most

$$\frac{M}{N}\binom{\beta\sqrt{N}+2}{2}\left(1 + \sum_{\gamma=1}^{2T} 3\gamma \max_v P^\gamma(u,v)\right)\,.$$

The remaining two cases will add to the same bound, so effectively this replaces a $4\max\{A_T, A_T^*\}$ in the original theorem with the expression $2\left(1+\max_u \sum_{\gamma=1}^{2T} 3\gamma \max_v P^\gamma(u,v)\right)$. □

To simplify, note that if $\max_{u,v} P^j(u,v) \le c + d^j$ then

$$\begin{aligned} 1+\sum_{j=1}^{2T} 3j \max_{u,v} P^j(u,v) &\le 1+3d\sum_{j=1}^{2T} j d^{j-1} + 3c\sum_{j=1}^{2T} j \\ &\le 1+\frac{3d}{(1-d)^2} + 3cT(2T+1)\,. \end{aligned} \tag{5.4}$$

Acknowledgment. The authors thank S. Miller and R. Venkatesan for several helpful discussions and for the pointers to E. Teske's work on discrete logarithms.

References

[1] D. Aldous and J. Fill, "Reversible Markov Chains and Random walks on Graphs," Book in preparation; available at http://www.stat.berkeley.edu/∼aldous.

[2] R. Crandall and C. Pomerance, "Prime Numbers : a computational perspective," Springer Verlag, 2nd ed., 2005, XVI.

[3] J. Fill, "Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process," *The Annals of Applied Probability*, Vol. 1, 1991, pp. 62–87.

[4] J.F. Le Gall and J. Rosen, "The range of stable random walks," *Ann. Probab.*, Vol. 19, 1991, pp.650–705.

[5] J-H. Kim and R. Montenegro and P. Tetali, "Near Optimal Bounds for Collision in Pollard Rho for Discrete Log," *Proc. of the 48th Annual Symposium on Foundations of Computer Science* (FOCS 2007), 2007.

[6] R. Lyons and Y. Peres and O. Schramm, "Markov chain intersections and the loop-erased walk," *Ann. Inst. H. Poincaré Probab. Statist.*, Vol. 39, no. 5, 2003, pp. 779–791.

[7] M. Mihail, "Conductance and Convergence of Markov Chains-A Combinatorial Treatment of Expanders," *Proc. of the 30th Annual Symposium on Foundations of Computer Science*, 1989, pp. 526–531.

[8] S. Miller and R. Venkatesan, "Spectral Analysis of Pollard Rho Collisions," *Proc. of the 7th Algorithmic Number Theory Symposium* (ANTS VII); Springer LNCS Vol. 4076, 2006, pp. 573–581.

[9] S. Miller and R. Venkatesan, Personal communications, 2007.

[10] S. Pohlig and M. Hellman, "An improved algorithm for computing logarithms over GF(p) and its cryptographic significance," *IEEE Trans. Information Theory*, Vol. 24, 1978, pp. 106-110.

[11] J.M. Pollard, “A Monte Carlo method for factorization,” *BIT Nord. Tid. f. Inf.* Vol. 15, 1975, pp. 331–334.

[12] J.M. Pollard, “Monte Carlo methods for index computation (mod p),” *Mathematics of Computation* **32** (143) 1978, pp. 918–924.

[13] C. Pomerance, “Elementary thoughts on discrete logarithms,” to appear in the proceedings of an MSRI workshop, J. Buhler and P. Stevenhagen, eds. (Preprint available at http://www.math.dartmouth.edu/∼carlp)

[14] V. Shoup, “Lower bounds for discrete logarithms and related problems,” *Proc. Advances in Cryptology* - EUROCRYPT ’97, 1997; Springer LNCS Vol. 1233, pp. 256–266.

[15] A. Sinclair, “Improved bounds for mixing rates of Markov chains and multicommodity flow,” *Combinatorics, Probability and Computing*, vol. 1, no. 4, 1992, pp. 351–370.

[16] E. Teske, “Speeding up Pollard’s rho method for computing discrete logarithms,” *Proc. of the 3rd Algorithmic Number Theory Symposium* (ANTS III); Springer LNCS Vol. 1423, pp. 541–554.

[17] E. Teske, “Square-root algorithms for the discrete logarithm problem (a survey),” In *Public Key Cryptography and Computational Number Theory*, Walter de Gruyter, 2001, pp. 283-301.