Tel Aviv University
The Raymond and Beverly Sackler
Faculty of Exact Sciences
The Blavatnik School of Computer Science

Partition Exchange and Anonymous Communication on Sparse Graphs

Submitted as a partial fulfillment of the
requirements towards Master of Science degree
Tel Aviv University
School of Computer Science

by
Tomer Levinboim

Under the supervision of
Prof. Amnon Ta-Shma

July 2010

Abstract

Over the past few decades several anonymous communication protocols have been suggested and studied, among which most notable are Chaum's protocol [Cha79] and its derivatives which have been rigorously analyzed.

Previous solutions assume at least one of the following assumptions:

1. All players are active at all times.
2. The network's graph is complete.

These two assumptions do not fit well with most networks and in particular, the Internet.

One particular solution is that of [CKKL99]. Improving on a work of Rackoff and Simon [RS93], they show that under these two assumptions, a variant of Chaum's protocol provides anonymous communication within $O(log^2(N))$ steps on a network of size N. Their proof relies on the convergence rate of a Markov chain they call "Matching Exchange" which is already referred to in [RS93].

In this thesis we define and analyze a similar Markov chain (we call "Partition Exchange") and show that it converges even faster. Our proof can be viewed as an extension of [CKKL99].

We are optimistic that this analysis will aid the design of an efficient new anonymous communication protocol which would not require the above two assumptions.

Acknowledgement

To my wife.

Contents

Chapter 1

Anonymous Communication

The phrase "anonymous" can take several interpretations. (1) we could wish to hide the *content* of a sent message (this is sometimes called "confidentiality"). (2) we might want to have *senders and receivers* anonymity. And (3) we would like to have *unlinkability*, meaning that even if an adversary knows the set $\{a_1, \ldots, a_n\}$ of senders and the set $\{b_1, \ldots, b_n\}$ of receivers, he cannot link the senders to the receivers.

The attack model also has several variants. In the *passive* model the adversary is curious but honest, i.e., it eavesdrops on the communication links/nodes under its control, but it does not deviate from the protocol. We call such an adversary an *eavesdropper*. An *active* adversary might change, initiate or delete messages. Both the passive and the active adversaries can be *static* meaning that they determine the communication links under their control before the protocol begins, or *dynamic* meaning that they may acquire communication links during the execution of the protocol and based on the communication so far.

Finally, there are the network topology and the *cost* issues. The desired network topology is a sparse graph with a small diameter, like the Internet graph. Common cost functions are:

- *time delay* - The time it takes a message to reach its destination.
- *message overhead* - The number of messages transmitted in the protocol per send request.

Current solutions can be divided into three groups: (1) A trusted party solution (See the survey [DD06, Sect. 2]) (2) Heuristics solutions which do not provide a security proof (many of which are based on Chaum's seminal paper from 1979 [Cha79]). A survey of this work can be found in [DD06, Sect. 3]. (3) Rigorous proofs, among which we note: Buses [BD03], Crowds [RR98], DC-Nets protocol [Cha88],Rackoff and Simon [RS93] rigorous security proof for a variant

of Chaum's protocol and its improvement [CKKL99] and finally [GKŁ04] and [BFG$^+$10].

We consider only those protocols with a rigorous proof - namely, those noted in point (3).

All of these protocols fail to be applicable because either:

- They require the participation of all nodes at each stage of the protocol, leading to a high message overhead when the number of active players n is much smaller than the network size N.
- They are not suitable for general sparse graphs.
- The security they provide is only proportional to the path length (e.g., Crowds).

Chaum's protocol [Cha81] hides the content of a message and its destination using multiple layers of encryption and can be viewed as a reduction from the unlinkability problem to that of *traffic analysis.* In the traffic analysis problem, n senders generate n indistinguishable packets and route them through a network of size N. The adversary then tries to analyze the traffic in order to link between senders and their intended receivers. Chaum does not give a formal proof of this reduction, however in 2005 Camenisch and Lysyanskaya [CL05] defined and designed a provably secure onion routing scheme. Using their work Chaum's protocol is indeed a provable reduction from unlinkability to traffic analysis. We therefore concern ourselves only with the traffic analysis problem.

The following solutions meet all but one of the desired properties:

- Relying on a protocol related to that of Chaum, [GKŁ04] achieve unlinkability after $polylog(n)$ steps for graphs with mixing time $polylog(n)$. However, it requires that $n = \Theta(N)$ players participate at all times.
- Based on Chaum's protocol, [BFG$^+$10, BFTS04] provide unlinkability within $O(log(n))$ steps and do not require that all players play at all times. In addition, They provide unlinkability even if the distribution of the receivers is not uniform on all players. Indeed, in reality the a-priori distribution is far from uniform as for example, people tend to communicate more with those speaking their own language. However, it assumes a relaxed adversary that controls only a constant fraction of the links and a complete graph.

1.1 [CKKL99] and our research

Based on [RS93], [CKKL99] show that when $n = \Theta(N)$ players are active, unlinkability can be reached within $O(log^2(n))$ steps on a complete graph, against an adversary that controls all the links and a constant fraction of the nodes. They analyze the following Markov chain which they call "Matching Exchange" (also known as "Pair and Swap" by [RS93]):

Matching Exchange: Let [n] be a set of positions, each containing a single element. The transition to the next state is then defined as follows:

1. Generate a random partial matching of size $\Theta(n)$ on the n positions.
2. For each edge in the matching, swap the content of the paired positions with independent probability $\frac{1}{2}$.

It is easy to see that this Markov chain is ergodic and reversible and therefore has $U_{\mathbb{S}_n}$ as its stationary distribution. A well known measure of convergence to the stationary distribution is the mixing time (See section 3.1.1). Briefly, the mixing time relates to the norm 1 distance between the distribution of the Markov chain at a given time step and its stationary distribution. [CKKL99] show that mixing time to $\epsilon = \frac{1}{n}$ is $O(log(n))$ steps.

One of the sources of randomness used in the Matching Exchange protocol is the choice of matchings. Fixing the matchings yields a (matching) *switching network*. [CKLK01] consider the Matching Exchange process run with a fixed sequence of matchings (that is, using only the randomness in step (2) of Matching Exchange) and on the state space of all 0-1 sequences with $\frac{n}{2}$ zeros and $\frac{n}{2}$ ones. They show that for almost any possible choice of matchings, the distribution of the process after $\tau = O(log(n))$ steps is almost uniform on the modified state space. In conjunction with [RS93] this proves that unlinkability can be reached within $O(log^2(n))$ steps when $n = \Theta(N)$ players are active at all times and on a complete graph.

In Chapter 3 we define and analyze a chain similar to Matching Exchange which we call "Partition Exchange":

Partition Exchange: Let [n] be a set of positions each containing a single element. Let m be the number of bins. The transition to the next state is then defined as follows:

1. Partition the $[n]$ elements into the m bins by placing each position in an independently and uniformly selected bin.
2. For each bin, permute the content of its positions independently and uniformly at random.

The stationary distribution of this Markov chain is also $U_{\mathbb{S}_n}$ for the same reason as above. In chapter 3 we prove its mixing time is $O(log_{\frac{n}{m}}(n))$ for $m = O(\frac{n}{log(n)})$.

Now consider modifying Partition Exchange to work on a fixed sequence of partitions. We would like to extend our result in a similar manner as [CKLK01] - that is, to show that for almost any fixed sequence of $poly(log_{\frac{n}{m}}(n))$ partitions, the distribution of the modified Partition Exchange protocol is close to uniform (note that the state space is left unchanged). Currently however, we are unable to do so. Nevertheless, assuming it can be shown, we can provide an anonymous communication protocol that has the following properties:

- On sparse graphs with diameter D, unlinkability can be reached already within $poly(D \cdot log_{\frac{n}{m}}(n))$ steps, even when $n \ll N$. In particular, this applies to the Internet.

- The protocol can provide anonymity services *even for a single player* at the cost of moderate message overhead.
- It is routing oblivious in the sense that as long as messages arrive to their specified destinations on time, the physical routing itself can be chosen by the adversary.

To achieve this our protocol requires the usage of a shared bulletin board and shared randomness which is perhaps its main drawback.[1]

[1] Once having a bulletin board, shared randomness can be obtained e.g., by using a pseudorandom generator and assuming at least one of the active players is honest.

Chapter 2

Preliminaries

2.1 Tail Bounds

2.1.1 Chernoff

Theorem 2.1.1. *[MR95] (Theorem 4.1) Let $X_1, \ldots, X_n$ be independent Boolean random variables, $X = \sum_i X_i$, $\mu = \mathbb{E}[X]$. Then,*

$$Pr[X \geq (1+\delta)\mu] \leq [e^{\delta}/(1+\delta)^{(1+\delta)}]^{\mu}$$

The following simple version of the Chernoff bound will be useful in our proof.

Theorem 2.1.2. *[MR95] (Theorem 4.1) Let $X_1, \ldots, X_n$ be independent Boolean random variables, $X = \sum_i X_i$, $\mu = \mathbb{E}[X]$. Then,*

$$Pr[X \leq (1-\delta)\mu] \leq exp(-\frac{\delta^2\mu}{2})$$

The following bound is useful when the expectation μ is very large:

Corollary. *Let $X_1, \ldots, X_n$ be independent Boolean random variables, $X = \sum_i X_i$, $\mu = \mathbb{E}[X]$. Assume $L \ll \mu$, then*

$$Pr[X \leq L] \leq exp(-(\frac{\mu}{2} - L))$$

2.1.2 Martingales and Azuma

Definition 2.1.1. [MR95] A sequence of $Z_1, \ldots, Z_n$ is said to be a *martingale with respect to* another sequence $X_1, \ldots, X_n$ If for all $1 \leq i \leq n$,

- $\mathbb{E}[Z_i] < \infty$
- $\mathbb{E}[Z_i | X_1, \ldots, X_{i-1}] = Z_{i-1}$

Definition 2.1.2. [MR95] The random variables $Z_1, \ldots, Z_n$ defined with respect to the random variables $X_1, \ldots, X_n$ and a function f as follows:

$$Z_i = \mathbb{E}_{X_{i+1},\ldots,X_n}[f(X_1, \ldots, X_n)|X_1, \ldots, X_i]$$

(where $X_1, \ldots X_i$ are treated as random variables) form a martingale and are known as a *Doob martingale*.

Definition 2.1.3. [MR95] Let $Z_0, Z_1, \ldots$ be a martingale sequence such that for each k $|Z_k - Z_{k-1}| \leq c_k$ where c_k may depend on k. Then by *Azuma's inequality*, for all t and any $\lambda > 0$

$$Pr[|Z_t - Z_0| \geq \lambda] \leq 2exp(-\frac{\lambda^2}{2\sum_{k=1}^{t} c_k^2})$$

The following will be used in conjunction with Azuma's inequality:

Definition 2.1.4. [MR95] Let $F : D^r \to \mathbb{R}$ be a real valued function with r arguments from a domain D. F is said to satisfy the *Lipschitz condition* if for any $(x_1, \ldots, x_r) \in D$ and any $k \in [r]$ and $y \in D$:

$$|F(x_1, \ldots, x_{k-1}, x_k, x_{k+1}, \ldots, x_r) - F(x_1, \ldots, x_{k-1}, y, x_{k+1}, \ldots, x_r)| \leq c_L$$

for some constant $c_L > 0$

Chapter 3

Partition Exchange

We consider a process we call "Partition Exchange" where at each time step, we randomly partition n positions (each storing a single element) among m bins and proceed to randomly permute the content of the positions within each bin. We are motivated by the fact that a variant of Partition Exchange models the adversary's knowledge under an anonymous communication protocol we designed.

[CKKL99] consider a similar process they call "Matching Exchange" - At each time step, a random partial matching of size $\Theta(n)$ is selected on the n positions and for each pair in the matching we independently and uniformly at random select whether to swap the elements between the two positions or leave them as is. [CKKL99] prove that convergence to the uniform distribution is met already after $O(logn)$ steps. The similarity between the two processes arises if we regard each pair in Matching Exchange as falling into the same bin. Put differently, we may degenerate Partition Exchange into Matching Exchange if we set $m = \Theta(n)$ and permute only bins that contain two elements.

Our goal is to show that Partition Exchange rapidly converges after $O(log_{\frac{n}{m}} n)$ time steps[1] and, in addition, to extract specific constants for which this happens. Our proof uses the delayed path coupling strategy developed in [CKKL99] and is an adaptation of [CKKL99] to our more general case.

3.1 Preliminaries

We introduce a few definitions: Let $\mathbb{S}_n$ denote the set of permutations on $[n]$. Let $\Delta : \mathbb{S}_n \times \mathbb{S}_n \to \{0, \ldots, n-1\}$ be such that for all $\pi, \pi' \in \mathbb{S}_n$, $\Delta(\pi, \pi')$ is the minimum number of transpositions required to transform π into π'. Clearly, Δ is a metric.

We denote the set of possible partitions of n elements to m bins as $\mathbb{P}_{n,m} = [m]^n$. Let $P \in \mathbb{P}_{n,m}$. We denote by $\mathbb{S}(P)$ the set of permutations that can be

[1] $\frac{n}{m}$ being the expected size of each bin.

attained on the partition P. That is, $\sigma \in \mathbb{S}(P)$ if for all $i, j \in [n]$, $\sigma(i) = j$ implies that i and j reside in the same bin.

Let H be some set and suppose $h \in H$. We denote by U_H the uniform distribution over H and we denote by $U_H(h)$ the probability of selecting h according to U_H. By $x \sim U_H$ we mean that x is selected according to U_H.

Fact 3.1.1. *[BY] Fix $p \in [n]$ and let σ be a randomly chosen permutation from $\mathbb{S}_n$. Define the random variable X to be the length of the cycle p resides in according to σ, then X is distributed uniformly on $[n]$.*

3.1.1 Mixing Time

Let $\mathfrak{M}$ be a Markov chain over a finite state space $\mathbf{S}$ with transition matrix M and a stationary distribution μ. Let $\pi^{(0)}$ be the initial distribution at time 0, and $\pi^{(t)} = M^t \pi^{(0)}$ the distribution at time t. A standard measure of convergence is the *mixing time* defined as:

$$\tau_{\mathfrak{M}}(\epsilon) \stackrel{\text{def}}{=} \min \left\{ T : \forall_{\pi^{(0)}}, \forall_{t \geq T} \, \|\pi^{(t)} - \mu\|_1 \leq \epsilon \right\}$$

3.1.2 Coupling

Let $\left\{(\pi^{(t)}, {\pi_R}^{(t)})\right\}_{t \in \mathbb{N}}$ be a stochastic process over $\mathbf{S} \times \mathbf{S}$, developing according to a Markov chain $\mathfrak{N}$. $\left\{(\pi^{(t)}, {\pi_R}^{(t)})\right\}$ is called a *coupling*[Ald83] for $\mathfrak{M}$ if marginally, both $\pi^{(t)}$ and ${\pi_R}^{(t)}$ evolve according to $\mathfrak{M}$. That is, $\pi^{(t+1)} = M\pi^{(t)}$ and ${\pi_R}^{(t+1)} = M{\pi_R}^{(t)}$ where M is the transition matrix of $\mathfrak{M}$. (Notice that we allow dependencies between $\pi^{(t)}$ and ${\pi_R}^{(t)}$). The mixing time of $\mathfrak{M}$ may be bounded using the following lemma:

Lemma 3.1.2. *(the coupling lemma) Suppose there is a coupling $\left\{(\pi^{(t)}, {\pi_R}^{(t)})\right\}$ for $\mathfrak{M}$ and that ${\pi_R}^{(0)}$ is the stationary distribution of $\mathfrak{M}$. If for all $t \geq T$, $\Pr[\pi^{(t)} \neq {\pi_R}^{(t)}] \leq \epsilon$ then $\tau_{\mathfrak{M}}(\epsilon) \leq T$.*

3.1.3 Path Coupling

Let $\mathbf{S}_2 \subseteq \mathbf{S} \times \mathbf{S}$ be a symmetric relation whose transitive closure is $\mathbf{S} \times \mathbf{S}$. Define a function $d : \mathbf{S} \times \mathbf{S} \to \mathbb{N}$ as follows: for *adjacent* states $(\pi, \pi_R) \in \mathbf{S}_2$, let $d(\pi, \pi_R) = 1$. For an arbitrary $(\pi, \pi_R) \in \mathbf{S} \times \mathbf{S}$, $d(\pi, \pi_R)$ is the length of the shortest path from π to π_R via $\mathbf{S}_2$.

The path coupling construction [BD97] reduces the task of finding a coupling that works on *all* initial pair of states, to that of finding one that needs to work only on *adjacent* states. Formally,

Lemma 3.1.3. *(the path coupling lemma) Suppose there exists a coupling $\left\{(\pi^{(t)}, {\pi_R}^{(t)})\right\}$ for $\mathfrak{M}$, and a symmetric relation $\mathbf{S}_2 \subseteq \mathbf{S} \times \mathbf{S}$ whose transitive closure is $\mathbf{S} \times \mathbf{S}$, such that for some $\delta < 1$, $\mathbb{E}[d(\pi^{(t+1)}, {\pi_R}^{(t+1)})] < \delta$ for all $t \in \mathbb{N}$ and adjacent $(\pi^{(t)}, {\pi_R}^{(t)}) \in \mathbf{S}_2$. Define $d_{max} = \max_{\pi, \pi' \in \mathbf{S}} d(\pi, \pi')$. Then,*

$$\tau_{\mathfrak{M}}(\epsilon) \leq \lceil \log(d_{max}\epsilon^{-1}) / \log(\delta^{-1}) \rceil$$

3.1.4 Delayed Path Coupling

The following Lemma taken from [CKKL99] (Lemma 4.2) relates the mixing time of a Markov chain $\mathfrak{M}$ to the behavior of a t-step path coupling. Note that the resulting coupling may be non-Markovian.

Lemma 3.1.4. *(the delayed path coupling lemma) Let T be a positive integer and let $\left\{(\pi^{(t)}, {\pi_R}^{(t)})\right\}$ be a coupling for $\mathfrak{M}$ such that for all $t \in \mathbb{N}$ and every adjacent $(\pi^{(t\cdot T)}, {\pi_R}^{(t\cdot T)}) \in \mathbf{S}_2$ we get that $\mathbb{E}[d(\pi^{(t(T+1))}, {\pi_R}^{(t(T+1))})] < \delta$ for some real $\delta < 1$. Then,*

$$\tau_{\mathfrak{M}}(\epsilon) \leq T \cdot \lceil \log(d_{max}\epsilon^{-1}) / \log(\delta^{-1}) \rceil$$

3.2 The Process

Let $[n]$ be a set of positions, each storing a single element and let m be some positive integer. We consider a Markov chain $\mathfrak{M}$ we call *Partition Exchange* on the set $\mathbb{S}_n$ of permutations of $[n]$ whose transition is defined as follows:

1. Partition $[n]$ among m bins by placing each position in an independently and uniformly selected bin. Let P denote the resulting partition.
2. Output a uniformly selected $\sigma \in_R \mathbb{S}(P)$.

It is clear that this Markov chain is ergodic and reversible, and therefore $\mathfrak{M}$ has $U_{\mathbb{S}_n}$ as its stationary distribution. We are interested in its rate of convergence.

Theorem 3.2.1. *For any $\epsilon > 0$, $n \geq 10^{10}$ and $m \leq \frac{n}{d_0 ln(n)}$ for $d_0 \geq 24$*

$$\tau_{\mathfrak{M}}(\epsilon) \leq (20 log_{\frac{n}{10m}}(n/49) + 10) \cdot \frac{ln(n/\epsilon)}{ln(n) - 10}$$

We shall use delayed path coupling (Lemma 3.1.4) to prove Theorem 3.2.1. However, we conjecture that in fact:

Conjecture 3.2.1.

$$\tau_{\mathfrak{M}}(\epsilon) \leq 4 log_{\frac{n}{m}}(n/\epsilon)$$

3.3 The Analysis

We define a Markov chain $\mathfrak{N} = (\pi^{(t)}, \pi_R^{(t)})_{t=0}^{T}$ with initial $\Delta(\pi^{(0)}, \pi_R^{(0)}) = 1$. Since $\Delta(\pi^{(0)}, \pi_R^{(0)}) = 1$, there exist $i \neq j \in [n]$ such that $\pi_R^{(0)} = (i, j)\pi^{(0)}$. For each $t \in [T]$ we pick a partition $P^{(t)}$ and a permutation $\sigma^{(t)}$ according to $\mathfrak{M}$. Let $\bar{P} = (P^{(1)}, \ldots, P^{(T)})$ and $\bar{\sigma} = (\sigma^{(1)}, \ldots, \sigma^{(T)})$. This defines the progress of the left coordinate of $\mathfrak{N}$. We now turn to define $\bar{\sigma}_R = (\sigma_R^{(1)}, \ldots, \sigma_R^{(T)}) = \bar{\sigma}_R(\bar{P}, \bar{\sigma}, \{i, j\})$ which will determine how the right coordinate of $\mathfrak{N}$ evolves. We wish to do so such that the second coordinate is also a faithful copy of $\mathfrak{M}$ and such that the two coordinates tend to meet.

3.3.1 An Observation

Suppose $\pi^{(0)} = (i,j)\pi_R^{(0)}$. Clearly, if i and j fall into the same bin at time 1, we may set $\sigma_R^{(1)} = \sigma^{(1)}(i,j)$ to get $\pi^1 = \pi_R^{(1)}$. However, this has probability exactly $\frac{1}{m}$ to occur.

Suppose $(\alpha_1, \ldots, \alpha_k = i)$, $(\beta_1, \ldots, \beta_l = j)$ are the (distinct) cycles of i and j in $\sigma^{(1)}$. We show that there are $min(k,l)$ ways to set $\sigma_R^{(1)}$ such that (1) For each $z \in \{1, \ldots, min(k,l)\}$, we maintain $\Delta(\pi^{(1)}, \pi_R^{(1)}) = 1$ and $\pi_R^{(1)} = (\alpha_z, \beta_z)\pi^{(1)}$ and (2) keep the same cycle structure. (1) implies that at time step 2, the probability that none of the (α_z, β_z) pairs falls into the same bins is at most $(1 - \frac{1}{m})^{min(k,l)}$ and (2) hints that the marginal distributions will be identical.

To see how we can keep distance 1 in $min(k,l)$ ways we choose some $z \in \{1, \ldots, min(k,l)\}$ and set $\sigma_R^{(1)}$ by exchanging between the $(z-1)$ length prefixes of the two cycles in $\sigma^{(1)}$ (for $z = 1$ there is no exchange). That is, we swap the cycles $(\alpha_1, \ldots, \alpha_k = i)$, $(\beta_1, \ldots, \beta_l = j)$ in $\sigma^{(1)}$ with $(\beta_1, \ldots, \beta_{z-1}, \alpha_z \ldots, \alpha_k = i)$ $(\alpha_1, \ldots, \alpha_{z-1}, \beta_z, \ldots, \beta_l = j)$ in $\sigma_R^{(1)}$. It is not difficult to verify that $\pi_R^{(1)} = (\alpha_z, \beta_z)\pi^{(1)}$

3.3.2 The Delayed Path Coupling

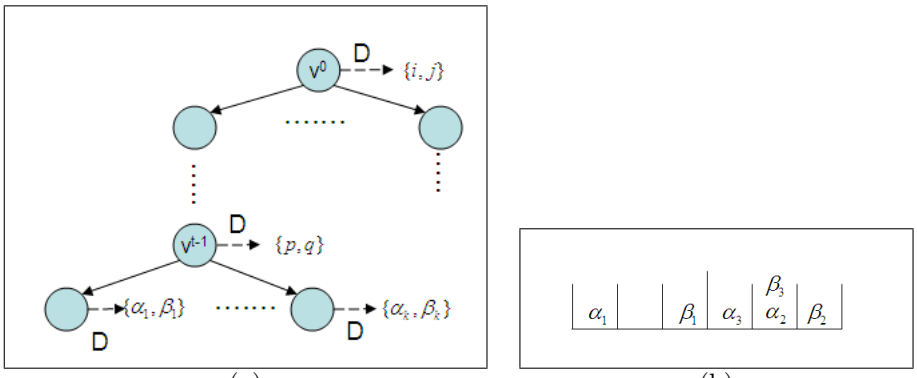
Definition 3.3.1. We define the *vertex labeled tree* $Tree^{(T)} = Tree^{(T)}(\bar{P}, \bar{\sigma}, \{i,j\})$ by induction on $t < T$. Initially $Tree^{(0)}$ contains a single vertex $v^{(0)}$ whose label $D(v^{(0)}) = \{i,j\}$ (A label is always a set of two distinct positions). Let $V^{(t-1)}$ denote the set of vertices at the $(t-1)$'th layer of the tree. A vertex $v \in V^{(t-1)}$ is called *isolated* at time t if it meets the following condition - There does not exist a vertex $u \neq v$ among the vertices of $V^{(t-1)}$ for which a position from $D(u)$ and a position from $D(v)$ reside in the same bin according to $P^{(t)}$ (See Figure 3.1.(b)).

Now, $Tree^{(t)}$ is defined based on $Tree^{(t-1)}$ as follows:

- Step 1: Set $W^{(t)} = \emptyset$ and let $I^{(t-1)} \subseteq V^{(t-1)}$ be the set of isolated vertices according to $P^{(t)}$. For each vertex $v \in I^{(t-1)}$ with $D(v) = \{p,q\}$, suppose $(\alpha_1, ..., \alpha_k = p)$, $(\beta_1, ..., \beta_l = q)$ are the cycles of p and q according to $\sigma^{(t)}$, then, for each $z \in \{1, \ldots, min(k,l)\}$ generate a vertex $v_z^{(t)}$ with label $D(v_z^{(t)}) = \{\alpha_z, \beta_z\}$ and add it to $W^{(t)}$.
- Step 2: If $|W^{(t)}| \leq |V^{(t-1)}|$
  - Discard $W^{(t)}$ and its defined labeling and for each $v \in V^{(t-1)}$ with $D(v) = \{p,q\}$, add a single vertex labeled $\{\sigma^{(t)}(p), \sigma^{(t)}(q)\}$ to $V^{(t)}$.

  Otherwise, we manage the growth of $V^{(t)}$ as follows:

  1. If $\frac{m}{36} < |W^{(t)}| < c_0 n$ for some constant parameters c_0 to be define later, then $V^{(t)}$ takes only the lexicographically first (according to the labels) $\frac{m}{36}$ vertices in $W^{(t)}$ with their defined labeling and we discard the rest.



(a) (b)

Figure 3.1: **(a)** The labeled tree at time t, where the positions p and q are in the label of $v^{(t-1)} \in V^{(t-1)}$ and according to $\sigma^{(t)}$, p and q reside in the cycles $(\alpha_1, \ldots, \alpha_k = p)$ and $(\beta_1, \ldots, \beta_l = q)$ (resp.) for $k = min(k, l)$.
(b) For $m = 6$, suppose $V^{(1)} = \{v_1, v_2, v_3\}$ with $D(v_1) = \{\alpha_1, \beta_1\}$, $D(v_2) = \{\alpha_2, \beta_2\}$, $D(v_3) = \{\alpha_3, \beta_3\}$. A possible partition $P^{(2)}$ is depicted where v_1 is an isolated vertex according to $P^{(2)}$, while v_2, v_3 are not.

2. Otherwise, set $V^{(t)} = W^{(t)}$ with its defined labeling.

We call a time step t for which we discard the intermediate layer $W^{(t)}$ *non-generating*.

Note that by the first condition in Step 2, the number of vertices in a layer is non-decreasing with respect to its former layer. Clearly, the labels of vertices generated by the same vertex are disjoint. Moreover, since we only use isolated vertices it is easy to see that for any specific layer, a position may appear in the label of at most one vertex in that layer. It is also easy to see that this remains true for non-generating time steps as well. We may conclude that the labels in a given layer are disjoint which implies that there can be at most $\frac{n}{2}$ vertices in any layer.

Definition 3.3.2. We say $(\bar{P}, \bar{\sigma}, \{i, j\})$ is *good* if there exists some $t \in [T]$ and $w \in V^{(t-1)}$ with label $D(w) = \{p, q\}$ such that both p and q fall into the same bin according to $P^{(t)}$. We call the first t for which such a w exists, the *good time step* of $(\bar{P}, \bar{\sigma}, \{i, j\})$, and we call the lexicographically first such $w \in V^{(t-1)}$, the *good vertex* of $(\bar{P}, \bar{\sigma}, \{i, j\})$.

The delayed path coupling construction: If $(\bar{P}, \bar{\sigma}, \{i, j\})$ is not good we set $\bar{\sigma}_R(\bar{P}, \bar{\sigma}, \{i, j\}) = \bar{\sigma}_R = \bar{\sigma}$. Otherwise, let (t, w) be the good time step and good vertex of $(\bar{P}, \bar{\sigma}, \{i, j\})$ and let $(w^{(0)} = v^{(0)}, w^{(1)}, \ldots, w^{(t-1)} = w)$ be the path of w in $Tree^{(T)}(\bar{P}, \bar{\sigma}, \{i, j\})$. Now,

- For all $\tau = 1, \ldots, t-1$: Set $\sigma_R^{(\tau)} = D(w^{(\tau)})\sigma^{(\tau)}D(w^{(\tau-1)})$.

- For t: Set $\sigma_R^{(t)} = \sigma^{(t)} D(w) = \sigma^{(t)} D(w^{(t-1)})$.
- For every $\tau \in \{t+1, \ldots, T\}$: Set $\sigma_R^{(t)} = \sigma^{(t)}$.

Thus $(\bar{\sigma}, \bar{\sigma}_R)$ define the T step progress of $\mathfrak{N}$. This completes the construction.

In the following subsections we prove the following Theorems:

- Theorem 3.4.1: $\mathfrak{N}$ is a delayed path coupling of $\mathfrak{M}$.
- Theorem 3.5.1 Let $i \neq j \in [n]$ and suppose $\pi_R^{(0)} = (i,j)\pi^{(0)}$. Suppose $m \leq \frac{n}{d_0 ln(n)}$ for $d_0 \geq 24$, then,

$$\mathbb{E}[\Delta(\pi^{(T)}, \pi_R^{(T)})] \leq e \cdot n^{-1/10}$$

For $T \geq 2log_{\frac{n}{10m}}(n) + 1$

Combining these two theorems with the delayed path coupling Lemma (Lemma 3.1.4) we prove Theorem 3.2.1:

Proof. Note that the metric Δ meets the condition of the delayed path coupling Lemma and that for all $(\pi, \pi') \in \mathbb{S} \times \mathbb{S}$, $\Delta(\pi, \pi') < n$. By Theorem 3.4.1, $\mathfrak{N}$ is a delayed path coupling of $\mathfrak{M}$ and since the Markov chain is time homogeneous, we may apply the delayed path coupling Lemma (Lemma 3.1.4). Specifically, for $\delta = e \cdot n^{-1/10}$ and $T = 2log_{\frac{n}{10m}}(n/49) + 1$ we obtain that:

$$\tau_{\mathfrak{M}}(\epsilon) \leq (2log_{\frac{n}{10m}}(n) + 1)\frac{ln(n\epsilon^{-1})}{ln(n^{1/10}/e)} \leq (20log_{\frac{n}{10m}}(n) + 10)\frac{ln(n\epsilon^{-1})}{ln(n) - 10}$$

as stated. □

3.4 $\mathfrak{N}$ is a Delayed Path Coupling of $\mathfrak{M}$

Let $\mathbb{H} = \left\{(\bar{P}, \bar{\sigma}) \in \mathbb{P}_{n,m}^T \times \mathbb{S}_n^T | \forall t \in [T], \sigma^{(t)} \in \mathbb{S}(P^{(t)})\right\}$ - That is $\mathbb{H}$ is the set of all T length sequences of partitions and T length sequences of permutations such that each permutation can be attained on its corresponding partition. In addition, let $\mathbb{H}_{\widetilde{\sigma}} = \left\{(\bar{P}, \bar{\sigma}) \in \mathbb{H} | \prod_t \sigma^{(t)} = \widetilde{\sigma}\right\}$ and let $N_2 = \{\{p,q\} | p, q \in [n], p \neq q\}$ We define a function $\Phi : \mathbb{H} \times N_2 \to \mathbb{P}_{n,m}^T \times \mathbb{S}_n^T$ and a function Φ_σ which is identical to Φ, but outputs only the sequence of permutations (i.e., if $\Phi(\bar{P}, \bar{\sigma}, \{i,j\}) = (\bar{\Pi}, \bar{\beta})$ then $\Phi_\sigma(\bar{P}, \bar{\sigma}, \{i,j\}) = \bar{\beta}$).

We shall show that Φ and Φ_σ have the following properties:

1. (Validness:) $\Phi(\bar{P}, \bar{\sigma}, \{i,j\}) \in \mathbb{H}$ for all $(\bar{P}, \bar{\sigma}) \in \mathbb{H}$ and $i, j \in [n]$
2. (Φ_σ represents our construction:) $\bar{\sigma}_R(\bar{P}, \bar{\sigma}, \{i,j\}) = \Phi_\sigma(\bar{P}, \bar{\sigma}, \{i,j\})$
3. (Equiprobability) Fix $i, j \in [n]$. For all $(\bar{\Pi}, \bar{\beta}) \in \mathbb{H}$,

$$Pr_{(\bar{P}, \bar{\sigma}) \sim U_{\mathbb{H}}}[\Phi(\bar{P}, \bar{\sigma}, \{i,j\}) = (\bar{\Pi}, \bar{\beta})] = U_{\mathbb{H}}(\bar{\Pi}, \bar{\beta})$$

Given the above properties of our construction we may now prove that $\mathfrak{N}$ is a T step delayed path coupling of $\mathfrak{M}$.

Theorem 3.4.1. *For all $\widetilde{\sigma} \in \mathbb{S}_n$*

$$Pr_{(\bar{P},\bar{\sigma})\sim U_{\mathbb{H}}}[\sigma^{(T)} \dots \sigma^{(1)} = \widetilde{\sigma}] = Pr_{(\bar{P},\bar{\sigma})\sim U_{\mathbb{H}}}[\sigma_R^{(T)} \dots \sigma_R^{(1)} = \widetilde{\sigma}]$$

and therefore $\mathfrak{N}$ is a coupling of $\mathfrak{M}$.

Proof.

$$\begin{aligned}
&Pr_{(\bar{P},\bar{\sigma})\sim U_{\mathbb{H}}}[\sigma_R^{(T)} \dots \sigma_R^{(1)} = \widetilde{\sigma}] \\
&= \sum_{\bar{\beta}\in\mathbb{S}_n^T:\prod_t \beta^{(t)}=\widetilde{\sigma}} Pr_{(\bar{P},\bar{\sigma})\sim U_{\mathbb{H}}}[\bar{\sigma}_R(\bar{P},\bar{\sigma},\{i,j\}) = \bar{\beta}] \\
&= \sum_{\bar{\beta}\in\mathbb{S}_n^T:\prod_t \beta^{(t)}=\widetilde{\sigma}} Pr_{(\bar{P},\bar{\sigma})\sim U_{\mathbb{H}}}[\Phi_\sigma(\bar{P},\bar{\sigma},\{i,j\}) = \bar{\beta}] \text{ (By property 2)} \\
&= \sum_{(\bar{\Pi},\bar{\beta})\in H_{\widetilde{\sigma}}} Pr_{(\bar{P},\bar{\sigma})\sim U_{\mathbb{H}}}[\Phi(\bar{P},\bar{\sigma},\{i,j\}) = (\bar{\Pi},\bar{\beta})] \\
&= \sum_{(\bar{\Pi},\bar{\beta})\in H_{\widetilde{\sigma}}} U_{\mathbb{H}}(\bar{\Pi},\bar{\beta}) \text{ (By property 3)} \\
&= \sum_{\bar{\beta}\in\mathbb{S}_n^T:\prod_t \beta^{(t)}=\widetilde{\sigma}} Pr_{(\bar{P},\bar{\sigma})\sim U_{\mathbb{H}}}[\bar{\sigma} = \bar{\beta}] = Pr_{(\bar{P},\bar{\sigma})\sim U_{\mathbb{H}}}[\sigma^{(T)} \dots \sigma^{(1)} = \widetilde{\sigma}]
\end{aligned}$$

as required. □

3.4.1 The Function Φ

We begin by defining the function Φ. Let $((P_\Phi^{(1)},\dots,P_\Phi^{(T)}),(\sigma_\Phi^{(1)},\dots,\sigma_\Phi^{(T)}))$ denote its output. Now, if $(\bar{P},\bar{\sigma},\{i,j\})$ is not good we simply set

$$((P_\Phi^{(1)},\dots,P_\Phi^{(T)}),(\sigma_\Phi^{(1)},\dots,\sigma_\Phi^{(T)})) = (\bar{P},\bar{\sigma})$$

Otherwise, let t and $w \in V^{(t-1)}$ be its good time step and good vertex and suppose the path from the root to w in $Tree^{(T)} = Tree(\bar{P},\bar{\sigma},\{i,j\})$ is $(w^{(0)} = v^{(0)}, w^{(1)},\dots,w^{(t-1)} = w)$. Then, for each $\tau \in [T]$:

- For all $1 \le \tau < t$: Suppose $D(w^{(\tau-1)}) = \{p,q\}$ and $(\alpha_1,\dots,\alpha_k = p)$ and $(\beta_1,\dots,\beta_l = q)$ are the cycles of p and q according to $\sigma^{(\tau)}$. Therefore, by the construction of $Tree$, $D(w^{(\tau)}) = \{\alpha_z,\beta_z\}$ for some $z \in \{1,\dots,min(k,l)\}$[2]. Define $f^{(\tau)} : [n] \to [n]$ such that for all $y \in \{1,\dots,z-1\}$ it maps α_y to β_y and vice versa, and on all other elements it acts as the identity. We define:

[2]This is true even when τ is non-generating, in which case $D(w^{(\tau)}) = \{\sigma^{(\tau)}(p),\sigma^{(\tau)}(q)\} = \{\alpha_1,\beta_1\}$

  - $\forall \alpha \in [n]$, $P_\Phi^{(\tau)}(\alpha) = P^{(\tau)}(f^{(\tau)}(\alpha))$. Note that $f^{(\tau)}$ modifies only the bins of p and q, and that p and q always remain in their respective bins.
  - $\sigma_\Phi^{(\tau)} = D(w^{(\tau)})\sigma^{(\tau)}D(w^{(\tau-1)})$.

- At time step t: We set $P_\Phi^{(t)} = P^{(t)}$ and $\sigma_\Phi^{(t)} = \sigma^{(t)}D(w)$
- For all $t < \tau \leq T$: We set $(P_\Phi^{(\tau)}, \sigma_\Phi^{(\tau)}) = (P^{(\tau)}, \sigma^{(\tau)})$

This completes the definition of Φ which in turn defines the function Φ_σ as well. It is easy to see by mere comparison between the two resulting sequences of permutations that $\bar{\sigma}_R(\bar{P}, \bar{\sigma}, \{i,j\}) = \Phi_\sigma(\bar{P}, \bar{\sigma}, \{i,j\})$. This shows property 2. Note that the construction is oblivious to time steps being generating or non-generating, to the size of the layer and to the growth management in step 2 of the labeled tree's construction.

We now relate our construction to the initial observation stated in the previous section.

Claim 3.4.1. *Suppose $(\alpha_1, \ldots, \alpha_k = p)$, $(\beta_1, \ldots, \beta_l = q)$ are the (distinct) cycles of p and q in $\sigma^{(t)}$ for some $t \in [T]$ and let $z \in \{1, \ldots, min(k,l)\}$. Moreover, let σ_* be the permutation obtained from $\sigma^{(t)}$ by replacing the cycles of p and q with $(\beta_1, \ldots, \beta_{z-1}, \alpha_z \ldots, \alpha_k = p)$ and $(\alpha_1, \ldots, \alpha_{z-1}, \beta_z, \ldots, \beta_l = q)$ (where z=1 means no change occurred). Then, $\sigma_* = (\alpha_z, \beta_z)\sigma^{(t)}(p,q)$*

Proof. We consider two cases:

- If $z = 1$ (such as when t is non-generating) then $\sigma_* = \sigma^{(t)}$ since no positions are swapped. Since $\sigma^{(t)}$ maps p to α_1 and q to β_1 it follows by inspection that $(\alpha_1, \beta_1)\sigma^{(t)}(p,q) = \sigma^{(t)}$ as well.
- If $z \neq 1$, the difference between $\sigma^{(t)}$ and $(\alpha_1, \beta_1)\sigma^{(t)}(p,q)$ is only in four positions - Namely, $\alpha_{z-1}, \beta_{z-1}, p$ and q. It is easy to verify by inspection that this is also the case with $\sigma^{(t)}$ and σ_*. Now, for α_{z-1}, since $z - 1 < min(k,l)$ we have that $(\alpha_z, \beta_z)(\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_l)(\alpha_{z-1}) = \beta_z$ and by inspection of the right hand side, it as well maps α_{z-1} to β_z. The other three cases can be shown identically.

Note that this shows $\sigma^{(t)}$ and $(\alpha_z, \beta_z)\sigma^{(t)}(p,q)$ have the same cycle decomposition up to renumbering of the positions. □

We now show that the construction is valid (property 1):

Claim 3.4.2. *For all $\tau \in [T]$, $\sigma_\Phi^{(\tau)} \in \mathbb{S}(P_\Phi^{(\tau)})$*

Proof. The claim is trivial for non-good $(\bar{P}, \bar{\sigma}, \{i,j\})$. Suppose $(\bar{P}, \bar{\sigma}, \{i,j\})$ is good and let t and w be its good time step and good vertex. Let $(w^{(0)} = v^{(0)}, \ldots, w^{(t-1)} = w)$ be the path from the root $v^{(0)}$ to w on $Tree(\bar{P}, \bar{\sigma}, \{i,j\})$. Now,

- For $1 \leq \tau \leq t-1$: Suppose $D(w^{(\tau-1)}) = \{p, q\}$ and $D(w^{(\tau)}) = \{\alpha_z, \beta_z\}$ as in the definition of Φ. Clearly:

$$\sigma_{\Phi}^{(\tau)} \triangleq D(w^{(\tau)})\sigma^{(\tau)}D(w^{(\tau-1)}) = (\alpha_z, \beta_z)\sigma^{(\tau)}(p, q)$$

  When $z = 1$ (such as in non-generating time steps) we have that $\sigma^{(\tau)} = \sigma_{\Phi}^{(\tau)}$ and since $f^{\tau} = id$ in this case, $P^{(\tau)} = P_{\Phi}^{(\tau)}$ and so the claim follows. Otherwise for $z \neq 1$, Claim 3.4.1 implies that $\sigma_{\Phi}^{(\tau)}$ was obtained from $\sigma^{(\tau)}$ by exchanging the the cycles of p and q with $(\beta_1, \ldots, \beta_{z-1}, \alpha_z \ldots, \alpha_k = p)$ and $(\alpha_1, \ldots, \alpha_{z-1}, \beta_z, \ldots, \beta_l = q)$. This fits exactly with $f^{(\tau)}$ since it as well exchanges the bin location of $(\alpha_1, \ldots, \alpha_{z-1})$ and $(\beta_1, \ldots, \beta_{z-1})$. Therefore, $\sigma_{\Phi}^{(\tau)} \in \mathbb{S}(P_{\Phi}^{(\tau)})$ as required.

- At time t: By the construction $P_{\Phi}^{(t)} = P^{(t)}$, $\sigma_{\Phi}^{(t)} = \sigma^{(t)}D(w)$ and by the definition of a good vertex, both the positions in $D(w)$ reside in the same bin. It is therefore easy to see that $\sigma_{\Phi}^{(t)} \in \mathbb{S}(P_{\Phi}^{(t)})$.

- The claim is trivial for all $t < \tau \leq T$.

□

Therefore the construction is indeed valid and property 1 holds. Our next goal is property 3 and for that we first prove:

Claim 3.4.3. *Fix $(\bar{P}, \bar{\sigma}) \in \mathbb{H}$ and $i, j \in [n]$ and denote $(\bar{P}_R, \bar{\sigma}_R) = \Phi(\bar{P}, \bar{\sigma}, \{i, j\})$. Suppose $(\bar{P}, \bar{\sigma}, \{i, j\})$ is good and let (t, w) be its good time step and good vertex. Then*

$$Tree^{(t-1)}(\bar{P}, \bar{\sigma}, \{i, j\}) = Tree^{(t-1)}(\bar{P}_R, \bar{\sigma}_R, \{i, j\})$$

With identical labeling of the vertices.

Proof. Generally, the growth of a tree at time step $\tau \in [T]$ is completely determined by the following factors: (1) the vertices at the $(\tau-1)$th layer (2) the partition to bins at time τ which determines which of the vertices are isolated (3) the permutation at time τ which determines the number of vertices each isolated vertex adds to $W^{(\tau)}$ and their labeling, and (4) the size of $W^{(\tau)}$.

Let $(w^{(0)} = v^{(0)}, w^{(1)}, \ldots, w^{(t-1)} = w)$ be the path from the root to w in $Tree^{(t-1)}(\bar{P}, \bar{\sigma}, \{i, j\})$. We prove the claim by induction on τ: Initially $v^{(0)}$ is the only vertex of both $Tree^{(0)}(\bar{P}, \bar{\sigma}, \{i, j\}) = Tree^{(0)}(\bar{P}_R, \bar{\sigma}_R, \{i, j\})$ with label $D(v^{(0)}) = \{i, j\}$ according to both. Let $1 \leq \tau \leq t-1$ and assume that the trees are equal up to and including the $(\tau-1)$th layer, with identical labeling of their vertices. We would like to show this property holds for τ. We consider two cases:

1. If τ is a non-generating time step then, as shown in the proof of Claim 3.4.2 for $z = 1$:

$$\sigma_R^{(\tau)} = \sigma^{(\tau)} \text{ and } P_R^{(\tau)} = P^{(\tau)}$$

Now, by the induction hypothesis, $V^{(\tau-1)} = V_R^{(\tau-1)}$ with identical labeling of the vertices and since the partitions are identical it follows that both sides have the same set of isolated vertices. Combined with $\sigma_R^{(\tau)} = \sigma^{(\tau)}$ we see that for both trees the same $W^{(\tau)}$ with identical labeling of its vertices will be created in step 1 of the construction. Since τ is non-generating $W^{(\tau)}$ will be discarded under both constructions and the same set of vertices with identical labeling will be created in both sides.

2. Otherwise τ is a generating time step. By the induction hypothesis $V^{(\tau-1)} = V_R^{(\tau-1)}$ with identical labeling of the vertices. Now, suppose $D(w^{(\tau-1)}) = \{p, q\}$. Since $w^{(\tau-1)}$ is isolated according to $P^{(\tau)}$ (it must be so, since only isolated vertices may have children in generating time steps and $w^{(\tau)}$ is its child) it follows that the positions appearing in the labels of all other vertices in $V^{(\tau-1)}$ do not reside in the same bin with neither p nor q. Now, since only the bins of p and q might be modified in $P_R^{(\tau)}$ under $f^{(\tau)}$, it follows that for all $w^{(\tau-1)} \neq v \in V^{(\tau-1)}$, v is isolated according to $P_R^{(\tau)}$ if and only if it is isolated according to $P^{(\tau)}$. Moreover, since only the cycles of p and q are modified in $\sigma_R^{(\tau)}$ (By Claim 3.4.1) it follows that such an isolated v adds the same number of vertices with identical labeling to $W^{(t)}$.

   The remaining vertex to be considered is $w^{(\tau-1)}$. By Claim 3.4.1 $\sigma_R^{(\tau)}$ is obtained by exchanging the cycles of p and q with $(\beta_1, \ldots, \beta_{z-1}, \alpha_z, \ldots, \alpha_k = p)$ and $(\alpha_1, \ldots, \alpha_{z-1}, \beta_z, \ldots, \beta_l = q)$ for some $z \in \{1, \ldots, min(k, l)\}$. Note that for any such z, p and q themselves remain in their distinct bins and therefore $w^{(\tau-1)}$ remains isolated under $P_R^{(\tau)}$. It follows by inspection of the two pairs of cycles that under both $\sigma^{(\tau)}$ and $\sigma_R^{(\tau)}$, $w^{(\tau-1)}$ adds $min(k, l)$ vertices labeled $\{\alpha_1, \beta_1\}, \{\alpha_2, \beta_2\}, \ldots$ to $W^{(t)}$.

   Finally, by the above argument it follows that the intermediate layer $W^{(t)}$ is identical in both constructions. Therefore, whatever takes place in step 2 of the construction holds for both trees and therefore $V_R^{(\tau)} = V^{(\tau)}$

We conclude that $Tree^{(\tau)}(\bar{P}, \bar{\sigma}, \{i, j\}) = Tree^{(\tau)}(\bar{P}_R, \bar{\sigma}_R, \{i, j\})$ with identical labeling. This proof holds for all $\tau < t$ and in particular for $t - 1$ as required.

□

Claim 3.4.4. *Fix $i, j \in [n]$. For all $(\bar{P}, \bar{\sigma}) \in \mathbb{H}$, $(\bar{P}, \bar{\sigma}, \{i, j\})$ is good if and only if $(\bar{P}_R, \bar{\sigma}_R, \{i, j\})$ is good. Moreover, if t and w are the good time step and good vertex of $(\bar{P}, \bar{\sigma}, \{i, j\})$ then t and w are the good time step and good vertex of $(\bar{P}_R, \bar{\sigma}_R, \{i, j\})$.*

Proof. If $(\bar{P}, \bar{\sigma}, \{i, j\})$ is not good then $(\bar{P}, \bar{\sigma}) = (\bar{P}_R, \bar{\sigma}_R)$ and the claim follows. Otherwise, suppose $(\bar{P}, \bar{\sigma}, \{i, j\})$ is good with t and $w \in V^{(t-1)}$ as its good time step and good vertex. Claim 3.4.3 ensures that the path from the root to w is identical in both $Tree^{(t-1)}(\bar{P}, \bar{\sigma}, \{i, j\})$ and $Tree^{(t-1)}(\bar{P}_R, \bar{\sigma}_R, \{i, j\})$ with

identical labeling along the path. Denote $(w^{(0)} = v^{(0)}, w^{(1)}, \ldots, w^{(\tau-1)} = w)$ as the path of w in the identical subtrees. We first show that none of the first $t-1$ steps can be a good time step according to $(\bar{P}_R, \bar{\sigma}_R, \{i,j\})$. We then show that t is indeed its good time step with $w \in V^{(t-1)}$ as the lexicographically first vertex for which the positions in its label fall into the same bin.

- For all $1 \leq \tau \leq t-1$, as explained before $w^{\tau-1}$ must be isolated according to $P^{(\tau)}$. Suppose $D(w^{\tau-1}) = \{p, q\}$, then by the definition of f^τ the only bins whose content may be changed in $P_R^{(\tau)}$ is that of p and q, however, p and q themselves reside in the same two distinct bins according to both $P^{(\tau)}$ and $P_R^{(\tau)}$ and therefore $w^{\tau-1}$ cannot be a good vertex at time τ. Moreover, since f^τ leaves all other bins as is, none of the other vertices can be a good vertex at time τ as well.
- By the construction $P^{(t)} = P_R^{(t)}$ and by Claim 3.4.3, $V^{(t-1)} = V_R^{(t-1)}$ with identical labeling of the vertices. It follows that w is also the lexicographically first vertex according to $P_R^{(t)}$ for which the positions in its label fall into the same bin. Since t is the first time step for which this occurs, it follows that t and w are the good time step and good vertex of $(\bar{P}_R, \bar{\sigma}_R, \{i,j\})$ as well. This also implies that $(\bar{P}_R, \bar{\sigma}_R, \{i,j\})$ is good.

□

Denoting $\Phi_{i,j}$ as the function Φ with a fixed third parameter $\{i,j\} \subseteq [n]$, we now use the two claims above to show that $\Phi_{i,j}$ is a bijection and $\Phi_{i,j}^2 = id$.

Claim 3.4.5. *Let $i, j \in [n]$, then $\Phi_{i,j}(\Phi_{i,j}(\bar{P}, \bar{\sigma}))) = (\bar{P}, \bar{\sigma})$ and therefore $\Phi_{i,j}$ is a bijection.*

Proof. If $(\bar{P}, \bar{\sigma}, \{i,j\})$ is not good then $\Phi_{i,j}(\bar{P}, \bar{\sigma}) = (\bar{P}, \bar{\sigma})$ and indeed $\Phi_{i,j}(\Phi_{i,j}(\bar{P}, \bar{\sigma}))) = (\bar{P}, \bar{\sigma})$. Otherwise, let t and w be its good time step and good vertex. By Claim 3.4.4, t and w are also the good time step and good vertex of $(\bar{P}_R, \bar{\sigma}_R, \{i,j\})$. Moreover, Claim 3.4.3 ensures that the path from the root to w is identical in both trees with identical labeling along the path. Let $(w^{(0)} = v^{(0)}, w^{(1)}, \ldots, w^{(\tau-1)} = w)$ be that path and denote $\Phi_{i,j}(\bar{P}_R, \bar{\sigma}_R) = (\bar{P}_*, \bar{\sigma}_*)$. Then, by Φ's definition:

- For all $\tau < t$:

$$\begin{aligned} \sigma_*^{(\tau)} &= D(w^{(\tau)})\sigma_R^{(\tau)}D(w^{(\tau-1)}) \\ &= D(w^{(\tau)})D(w^{(\tau)})\sigma^{(\tau)}D(w^{(\tau-1)})D(w^{(\tau-1)}) \\ &= \sigma^{(\tau)} \end{aligned}$$

  and $\forall \alpha \in [n]$:

$$\begin{aligned} P_*^{(\tau)}(\alpha) &= P_R^{(\tau)}(f^{(\tau)}(\alpha)) \\ &= P^{(\tau)}(f^{(\tau)}(f^{(\tau)}(\alpha))) \\ &= P^{(\tau)}(\alpha) \end{aligned}$$

- For t we have that $P_*^{(t)} = P_R^{(t)} = P^{(t)}$ and

$$\sigma_*^{(t)} = \sigma_R^{(t)} D(w) = \sigma_R^{(t)} D(w) D(w) = \sigma^{(t)}$$

- For all $t < \tau \leq T$, $(P_*^{(\tau)}, \sigma_*^{(\tau)}) = (P_R^{(\tau)}, \sigma_R^{(\tau)}) = (P^{(\tau)}, \sigma^{(\tau)})$

□

Finally, we turn to prove property 3.

Lemma 3.4.2. *Fix $i, j \in [n]$. For all $(\bar{\Pi}, \bar{\beta}) \in \mathbb{H}$,*

$$Pr_{(\bar{P},\bar{\sigma})\sim U_{\mathbb{H}}}[\Phi_{i,j}(\bar{P}, \bar{\sigma}) = (\bar{\Pi}, \bar{\beta})] = U_{\mathbb{H}}(\bar{\Pi}, \bar{\beta})$$

Proof. By Claim 3.4.5,

$$\begin{aligned} Pr_{(\bar{P},\bar{\sigma})\sim U_{\mathbb{H}}}[\Phi_{i,j}(\bar{P}, \bar{\sigma}) = (\bar{\Pi}, \bar{\beta})] &= Pr_{(\bar{P},\bar{\sigma})\sim U_{\mathbb{H}}}[\Phi_{i,j}(\Phi_{i,j}(\bar{P}, \bar{\sigma})) = \Phi_{i,j}(\bar{\Pi}, \bar{\beta})] \\ &= Pr_{(\bar{P},\bar{\sigma})\sim U_{\mathbb{H}}}[(\bar{P}, \bar{\sigma}) = \Phi_{i,j}(\bar{\Pi}, \bar{\beta})] \end{aligned}$$

Where the first transition is valid since $\Phi_{i,j}$ is a bijection.

Now, if $(\bar{\Pi}, \bar{\beta}, \{i, j\})$ is not good then $\Phi_{i,j}(\bar{\Pi}, \bar{\beta}) = (\bar{\Pi}, \bar{\beta})$ and the claim is trivial. Otherwise $(\bar{\Pi}, \bar{\beta})$ is good. Denote $(\bar{\Pi}_R, \bar{\beta}_R) = \Phi_{i,j}(\bar{\Pi}, \bar{\beta})$. Note now that for all $\tau \in [T]$, $\Pi^{(\tau)}$ and $\Pi_R^{(\tau)}$ have the same structure in the sense that each bin contains the same number of elements. Therefore,

$$Pr_{\bar{P}\sim U_{\mathbb{P}_{n,m}^T}}[\bar{P} = \bar{\Pi}] = Pr_{\bar{P}\sim U_{\mathbb{P}_{n,m}^T}}[\bar{P} = \bar{\Pi}_R] \tag{3.1}$$

Now for all $\tau \in [T]$ given the fact that $P^{(\tau)}$ and $P_R^{(\tau)}$ have the same structure, $\mathbb{S}(P^{(\tau)})$ and $\mathbb{S}(P_R^{(\tau)})$ span sets of the same size and it follows that:

$$Pr[\bar{\sigma} = \bar{\beta} | \bar{P} = \bar{\Pi}] = Pr[\bar{\sigma} = \bar{\beta}_R | \bar{P} = \bar{\Pi}_R] \tag{3.2}$$

Combining 3.1 and 3.2 we get:

$$\begin{aligned} Pr_{(\bar{P},\bar{\sigma})\sim U_{\mathbb{H}}}[(\bar{P}, \bar{\sigma}) = (\bar{\Pi}_R, \bar{\beta}_R)] &= Pr_{(\bar{P},\bar{\sigma})\sim U_{\mathbb{H}}}[(\bar{P}, \bar{\sigma}) = (\bar{\Pi}, \bar{\beta})] \\ &= U_{\mathbb{H}}(\bar{\Pi}, \bar{\beta}) \end{aligned}$$

as required. □

3.5 Convergence Rate of $\mathfrak{M}$

Claim 3.5.1. *Let $i \neq j \in [n]$ and suppose $\pi_R^{(0)} = (i, j)\pi^{(0)}$. If $(\bar{P}, \bar{\sigma}, \{i, j\})$ is good then $\pi^{(T)} = \pi_R^{(T)}$, otherwise $\Delta(\pi^{(T)}, \pi_R^{(T)}) = 1$.*

Proof. If $(\bar{P}, \bar{\sigma}, \{i,j\})$ is not good, then $(\bar{P}_R, \bar{\sigma}_R) = (\bar{P}, \bar{\sigma})$ and it is clear that the distance remains 1. Otherwise, let t and w be the good time step and good vertex of $(\bar{P}, \bar{\sigma})$ and let $(w^{(0)} = v^{(0)}, w^1, \ldots, w^{(t-1)} = w)$ be the path of w in $Tree(\bar{P}, \bar{\sigma}, \{i,j\})$. Now,

$$\begin{aligned}\pi_R^{(T)} &= \sigma_R^{(T)} \ldots \sigma_R^{(1)} \pi_R^{(0)} \\ &= \sigma^{(T)} \ldots \sigma^{(t+1)} \sigma^{(t)} D(w) D(w^{(t-1)}) \sigma^{(t-1)} D(w^{(t-2)}) \ldots D(w^{(1)}) \sigma^{(1)} D(w^{(0)}) D(v^{(0)}) \pi^{(0)} \\ &= \sigma^{(T)} \ldots \sigma^{(1)} \pi^{(0)} = \pi^{(T)}\end{aligned}$$

as required. □

We now turn to prove Theorem 3.5.1.

Theorem 3.5.1. *Let $i \neq j \in [n]$ and suppose $\pi_R^{(0)} = (i,j)\pi^{(0)}$. Suppose $m \leq \frac{n}{d_0 ln(n)}$ for $d_0 \geq 24$, then,*

$$\mathbb{E}[\Delta(\pi^{(T)}, \pi_R^{(T)})] \leq e \cdot n^{-1/10}$$

For $T \geq 2log_{\frac{n}{10m}}(n) + 1$

Proof. Considering Claim 3.5.1:

$$\begin{aligned}\mathbb{E}[\Delta(\pi^{(T)}, \pi_R^{(T)})] &= Pr[\Delta(\pi^{(T)}, \pi_R^{(T)}) = 1] \\ &= Pr[(\bar{P}, \bar{\sigma}, \{i,j\}) \text{ is not good}] \\ &= Pr[\forall t \in [T], t \text{ is not a good time step of } (\bar{P}, \bar{\sigma}, \{i,j\})]\end{aligned}$$

Suppose we show that for some T_0, $|V^{(T_0)}| \geq c_0 n$ with high probability (c_0 as in the construction). Then the probability that none of vertices in $V^{(t-1)}$ are good according to $P^{(t)}$ for all $t = T_0 + 1, \ldots, T$ is at most:

$$exp(-\frac{c_0 n}{m}(T - T_0)) \tag{3.3}$$

We call a bin *balanced* at time t if according to $P^{(t)}$, it contains at least $\frac{2n}{3m}$ positions and in addition we call a partition *balanced* if all its bins are balanced. Since the positions are distributed independently and uniformly at random, we may apply Chernoff (Theorem 2.1.2) to conclude that a bin is not balanced with probability at most $exp(-\frac{n}{18m})$ and so using the union bound, a partition is not balanced with probability at most

$$m \cdot exp(-\frac{n}{18m}) \tag{3.4}$$

We state three major lemmas whose proof will be deferred. In these Lemmas we assume that $18ln(n) \leq \frac{m}{36}$.

1. **Growth to $\Omega(ln(n))$ vertices:**
   Lemma 3.6.1: Fix partitions $P^{(1)}, \ldots, P^{(D)}$ for $D = 0.9log_{\frac{n}{7.5m}}(n)$ and suppose they are balanced. Then, the labeled tree has $|V^{(D)}| < 18ln(n)$ with probability at most
   $$\frac{18 \cdot ln(n)}{n^{0.9/2}} + \frac{2D(18ln(n))^2}{m} \tag{3.5}$$

2. **Growth to $\Omega(m)$ vertices:**
   Lemma 3.6.2: Suppose $|V^{(D)}| \geq 18ln(n)$ for some integer $D > 0$. Then, after $D_1 = log_{\frac{n}{10m}}(\frac{m}{72ln(n)})$ steps the probability that $|V^{(D+D_1)}| < \frac{m}{36}$ is at most
   $$D_1(2n^{-\frac{18}{16}(2e^{-1/18}-1.5)^2} + m \cdot e^{-\frac{n}{18m}} + n^{-(0.05)^2\frac{n}{m}}) \tag{3.6}$$

3. **Growth to $\Theta(n)$ vertices:**
   Lemma 3.6.5: Suppose $|V^{(D+D_1)}| \geq \frac{m}{36}$ for some integers $D, D_1 > 0$. Then, the probability that $|V^{(D+D_1+1)}| < \frac{n}{342}$ is at most
   $$2n^{-\frac{18}{16}(2e^{-1/18}-1.5)^2} + m \cdot e^{-\frac{n}{18m}} + n^{-(0.05)^2\frac{n}{m}} \tag{3.7}$$

We condition on the first D partitions being balanced (Note that the other $D_1 + 1$ also turn out to be balanced, this conditioning happens within the Lemmas). In addition, we take $T_0 = D + D_1 + 1$ and note that:

$$\begin{aligned} T_0 + 2 &= 0.9log_{\frac{n}{7.5m}}(n) + log_{\frac{n}{10m}}(\frac{m}{72ln(n)}) + 3 \\ &< log_{\frac{n}{10m}}(n^{0.9}) + log_{\frac{n}{10m}}(\frac{m}{72ln(n)} \cdot (\frac{n}{10m})^2) + 1 \\ &< 2log_{\frac{n}{10m}}(n/49) + 1 \end{aligned}$$

Taking $T \geq 2log_{\frac{n}{10m}}(n/49) + 1$ for our choice of $n \geq 10^{10}$, and assuming $m \leq \frac{n}{d_0 ln(n)}$ we obtain that the sum of the the error terms 3.3-3.7 is at most $e \cdot n^{-1/10}$. This yields:

$$\mathbb{E}[\Delta(\pi^{(T)}, \pi_R^{(T)})] \leq e \cdot n^{-1/10}$$

□

3.6 Proof Of Lemmas

In the next three sections we state and prove the lemmas concerning with the growth of the labeled tree. In section 3.6.1 we show that after some D steps, the Dth layer has $\Omega(ln(n))$ vertices with high probability. We do so by showing that the growth in the first layers is almost independent. In sections 3.6.2 and 3.6.3 we argue that having $\Omega(ln(n))$ enables us to show (using Chernoff) that with high probability, each subsequent layer is expected to grow by a factor of $\Omega(\frac{n}{m})$ where here we use the independence between different vertices that belong to the same layer. Finally, in section 3.6.4 we prove two auxiliary lemmas.

3.6.1 Growth to $\Omega(ln(n))$

Consider the labeled tree that would be generated if we were to change the following two things in the construction (1) all vertices, isolated or not, would generate vertices in the manner described and, (2) we do not impose any restriction on the size of the layer such as those in step 2 of the construction. We call the resulting labeled tree the *non-pruned* labeled tree. Notice that in this tree, the labels in a layer may not be disjoint, or even repeated in several distinct vertices. In other words, the size of a layer is not bounded. The following claim shall help us in our analysis.

Claim 3.6.1. *Fix partitions $P^{(1)}, \ldots, P^{(D)}$ for $D = 0.9log_{\frac{n}{7.5m}}(n)$ and suppose the partitions are balanced. Then, the non-pruned tree has $|V^{(D)}| < 18ln(n)$ with probability at most $\frac{18 \cdot ln(n)}{n^{0.9/2}}$.*

Proof. For all $t \in [D]$, let $X^{(t)} = \frac{|V^{(t)}|}{|V^{(t-1)}|}$ and $Y^{(t)} = ln(X^{(t)})$. In addition, let $Y = \sum_{t=1}^{D} Y^{(t)}$. Clearly $|V^{(D)}| = \prod_{t=1}^{D} X^{(t)} = exp(Y)$ and therefore, if $Y \geq ln(18ln(n))$ then $exp(Y) \geq 18ln(n)$ as we desire. We are therefore required to bound $Pr[Y < ln(18ln(n))]$ for our choice of D.

We now determine $\mathbb{E}[Y^{(t)}]$ for all $t \in [D]$. Let $v \in V^{(t-1)}$ with $D(v) = \{p, q\}$ and denote by $g(v)$ the number of vertices v generates according to $\sigma^{(t)}$. Now,

$$Y^{(t)} = ln(\frac{|V^{(t)}|}{|V^{(t-1)}|}) = ln(\frac{\sum_{v \in V^{(t-1)}} g(v)}{|V^{(t-1)}|}) \tag{3.8}$$

$$\geq \frac{1}{|V^{(t-1)}|} \cdot \sum_{v \in V^{(t-1)}} ln(g(v)) \tag{3.9}$$

by the convexity of the ln function. For all $v \in V^{(t-1)}$, $g(v) = min(k_p, k_q)$ where k_p and k_q are the cycle lengths of p and q (resp.) in $\sigma^{(t)}$. By Fact 3.1.1, k_p and k_q are distributed uniformly over their respective bin size and therefore since we assume $P^{(t)}$ is balanced, it follows by Equation 3.9 and Claim 3.6.2 below, that:

$$\mathbb{E}[Y^{(t)}] \geq \frac{1}{|V^{(t-1)}|} \cdot \sum_{v \in V^{(t-1)}} \mathbb{E}[ln(g(v))] \geq ln(\frac{n}{7.5m}) \tag{3.10}$$

Setting $D = 0.9ln(n)/ln(\frac{n}{7.5m})$ we obtain $\mathbb{E}[Y] \geq 0.9ln(n)$. Now, 3.10 is true for any history $Y^{(1)}, \ldots, Y^{(t-1)}$, therefore we may apply Chernoff (Corollary 2.1.1) to obtain the stated bound. □

The claim implies that with high probability the Dth layer of the non-pruned tree contains at least $18ln(n)$ vertices. Note that the regular labeled tree evolves identically to the non-pruned tree conditioned on all vertices being isolated. Let $D' \leq D$ denote the first time step for which the layer of the non-pruned tree contains at least $K = 18ln(n)$ vertices. By the construction, the size of all previous layers is at most K vertices and therefore for any $t \in [D' - 1]$, the probability that there is some non-isolated vertex in $|V^{(t-1)}|$according to

$P^{(t)}$ is at most $\binom{2K}{2}\frac{1}{m} < \frac{2(18ln(n))^2}{m}$. Since $D' \leq D$ we can upper bound the probability of the existence of a non-isolated vertex in any of the first $D'-1$ layers by $\frac{2D(18ln(n))^2}{m}$. Note that due to the modifications in the growth of the non-pruned tree, we require that $18ln(n) < \frac{m}{36}$. This results with the following Lemma:

Lemma 3.6.1. *Fix partitions $P^{(1)}, \ldots, P^{(D)}$ for $D = 0.9log_{\frac{n}{7.5m}}(n)$ and suppose they are balanced. Then, the labeled tree has $|V^{(D)}| < 18ln(n)$ with probability at most*

$$\frac{18 \cdot ln(n)}{n^{0.9/2}} + \frac{2D(18ln(n))^2}{m} \tag{3.11}$$

3.6.2 Growth to $\Omega(m)$

Lemma 3.6.2. *Suppose $|V^{(D)}| \geq 18ln(n)$ for some integer $D > 0$. Then, after $D_1 = log_{\frac{n}{10m}}(\frac{m}{72ln(n)})$ steps the probability that $|V^{(D+D_1)}| < \frac{m}{36}$ is at most*

$$D_1(2n^{-\frac{18}{16}(2e^{-1/18}-1.5)^2} + m \cdot e^{-\frac{n}{18m}} + n^{-(0.05)^2\frac{n}{m}}) \tag{3.12}$$

Proof. Consider the following Lemma whose proof (using Azuma's inequality) is deferred:

Lemma 3.6.3. *Suppose $|V^{(t)}| = r \leq \frac{m}{36}$ Then, the probability that less than $\frac{r}{2}$ vertices are isolated according to $P^{(t+1)}$ is upper bounded by $2exp(-\frac{(2e^{-1/18}-1.5)^2 r}{16})$*

note that the lemma remains true when using balanced partition (rather then a random partition from $\mathbb{P}_{n,m}$, with the additional error term of $m \cdot exp(\frac{-n}{18m})$. We condition on the partitions being balanced. Now, for any balanced partition, a similar argument as in the proof of Lemma 3.6.1 shows that each isolated vertex is expected to generate at least $\frac{2n}{9m}$ vertices. Given that the partitions are balanced, the number of vertices each isolated vertex generates is independent from the other isolated vertices in that layer and we may apply Chernoff (Theorem 2.1.2) to obtain the following Lemma:

Lemma 3.6.4. *Given any balanced partition, a set of s isolated vertices generates less than $0.95\frac{2ns}{9m}$ vertices with probability at most $exp(-\frac{(0.05)^2 ns}{9m})$*

Summing up the probabilities of failure we obtain that for all $t > D$, as long as $|V^{(t)}| \leq \frac{m}{36}$, less than $0.95\frac{ns}{9m}$ are added to $W^{(t+1)}$ with probability at most:

$$2exp(-\frac{(2e^{-1/18}-1.5)^2|V^{(t)}|}{16}) + m \cdot exp(-\frac{n}{18m}) + exp(-\frac{(0.05)^2 n|V^{(t)}|}{18m})$$

We remind the reader that by step 2 of the construction, when $\frac{m}{36} < W^{(t)} < c_0 n$ we set $V^{(t)}$ to be of size $\frac{m}{36}$. In any case, since there are at least $18ln(n)$ vertices in each of the subsequent layers, it follows that after $D_1 = log_{\frac{n}{10m}}(\frac{m}{72ln(n)})$ steps we have $|V^{(D)}| < \frac{m}{36}$ with probability at most

$$D_1(2n^{-\frac{18}{16}(2e^{-1/18}-1.5)^2} + m \cdot e^{-\frac{n}{18m}} + n^{-(0.05)^2\frac{n}{m}})$$

□

3.6.3 Growth to $\Theta(n)$

Lemma 3.6.5. *Suppose $|V^{(D+D_1)}| \geq \frac{m}{36}$ for some integers $D, D_1 > 0$. Then, the probability that $|V^{(D+D_1+1)}| < \frac{n}{342}$ is at most*

$$2n^{-\frac{18}{16}(2e^{-1/18}-1.5)^2} + m \cdot e^{-\frac{n}{18m}} + n^{-(0.05)^2\frac{n}{m}} \tag{3.13}$$

Proof. The bound follows immediately by applying Lemmas 3.6.3 and 3.6.4. Except for the error probability, we get that:

$$|V^{(D+D_1+1)}| \geq 0.95 \cdot \frac{2n}{9m} \cdot \frac{1}{2} \cdot \frac{m}{36} \geq \frac{n}{342}$$

□

3.6.4 The Expected Number of Isolated Vertices

Proof of Lemma 3.6.3. Let $U^{(t)} = \bigcup_{v \in V^{(t)}} D(v)$ be the set of positions appearing in some label belonging to a vertex from $V^{(t)}$. Recall that in any given layer, each position appears in the label of at most one vertex and therefore $|U^{(t)}| = 2r$ where $|V^{(t)}| = r$. Let $Z^{(t+1)}$ denote the number of positions from $U^{(t)}$ which according to $P^{(t+1)}$ reside with some other position from $U^{(t)}$. The probability that a specific position from $U^{(t)}$ does not reside with any other position from $U^{(t)}$ in its bin is $(1 - \frac{1}{m})^{2r-1} \geq e^{-\frac{2r}{m}} \geq e^{-1/18}$ and therefore $\mathbb{E}[Z^{(t+1)}] \leq 2r(1 - e^{-1/18})$.

Suppose we show that with high probability at most $\frac{r}{2}$ of the positions from $U^{(t)}$ reside with some other position from $U^{(t)}$ in their bin according to $P^{(t+1)}$. Conditioned on this event, these positions belong to at most $\frac{r}{2}$ vertices from $V^{(t)}$, which means that the rest $\frac{r}{2}$ must be isolated according to $P^{(t+1)}$ as we require. We therefore bound the probability that more than $\frac{r}{2}$ positions do not reside alone in their bin.

We observe the process of placing the positions in the bins and analyze it in a similar fashion as [MR95], Exercise 4.12. Let $X_1^{(t+1)}, \ldots, X_{2r}^{(t+1)}$ be random variables where for each $k \in [2r]$, $X_k^{(t+1)}$ denotes the index of the kth position in $U^{(t)}$ according to $P^{(t+1)}$. Notice that $X_1^{(t+1)}, \ldots, X_{2r}^{(t+1)}$ are independent. We can view $Z^{(t+1)}$ as a function $f(X_1^{(t+1)}, \ldots, X_{2r}^{(t+1)})$ that returns the number of positions sharing their bin with at least one other position. It is easy to see that moving any position from one bin to another may change f's value by at most 2. Therefore, f satisfies the Lipschitz condition with constant $c_L = 2$. We now define the sequence of random variables $Z_0^{(t+1)}, \ldots, Z_{2r}^{(t+1)}$ as follows:

1. $Z_0^{(t+1)} = \mathbb{E}_{X_1^{(t+1)}, \ldots, X_{2r}^{(t+1)}}[f(X_1^{(t+1)}, \ldots, X_{2r}^{(t+1)})]$
2. For all $k \in [2r]$, $Z_k^{(t+1)} = \mathbb{E}_{X_{k+1}^{(t+1)}, \ldots, X_{2r}^{(t+1)}}[f(X_1^{(t+1)}, \ldots, X_{2r}^{(t+1)}) | X_1^{(t+1)}, \ldots, X_k^{(t+1)}]$

note that $Z_0^{(t+1)}, \ldots, Z_{2r}^{(t+1)}$ form a Doob martingale (Definition 2.1.2). It is clear that for all $k \in [2r]$, $|Z_k^{(t+1)} - Z_{k-1}^{(t+1)}| \leq 2$. Consequently, by Azuma's inequality:

$$Pr[|Z_{2r}^{(t+1)} - Z_0^{(t+1)}| \geq \lambda] \leq 2exp(-\frac{\lambda^2}{16r})$$

For all $\lambda > 0$. This is exactly the bound we are looking for since $Z_0^{(t+1)} = \mathbb{E}[Z^{(t+1)}]$ and $Z_{2r}^{(t+1)} = Z^{(t+1)}$. Now

$$\begin{aligned} Pr[Z^{(t+1)} > 0.5r] &< Pr[|Z^{(t+1)} - \mathbb{E}[Z^{(t+1)}| > 0.5r - \mathbb{E}[Z^{(t+1)}]] \\ &< 2exp(-\frac{(0.5r - \mathbb{E}[Z^{(t+1)}])^2}{16r}) \\ &\leq 2exp(-\frac{(2e^{-1/18} - 1.5)^2 r}{16}) \end{aligned}$$

□

3.6.5 The Expected Growth Size

Claim 3.6.2. *Let K be some integer and suppose $i, j \sim U_K$, then*

1. $\mathbb{E}[min(i,j)] \geq \frac{K}{3}$ *and,*
2. $\mathbb{E}[ln(min(i,j))] \geq ln(K/5)$ *for* $K \geq 11$

Proof.

$$\begin{aligned} \mathbb{E}[min(i,j)] &= \frac{1}{K^2}\sum_{i,j=1}^{K} min(i,j) = \frac{1}{K^2}[2\sum_{1\leq i<j\leq K} i + \sum_{i=1}^{K} i] && (3.14) \\ &= \frac{1}{K^2}[2\sum_{i=1}^{K} i(K-i) + \sum_{i=1}^{K} i] && (3.15) \\ &= \frac{1}{K^2}[\frac{1}{2}(2K+1)(K+1)K - \frac{1}{3}K(K+1)(2K+1)] && (3.16) \\ &\geq \frac{K}{3} && (3.17) \end{aligned}$$

Using an identical analysis up until 3.15 above we obtain:

$$\begin{aligned} \mathbb{E}[ln(min(i,j))] &= \frac{1}{K^2}[(2K+1)\sum_{i=1}^{K} ln(i) - 2\sum_{i=1}^{K} i \cdot ln(i)] && (3.18) \\ &= \frac{1}{K^2}[(2K+1)ln(K!) - 2\sum_{i=1}^{K} i \cdot ln(i)] && (3.19) \end{aligned}$$

Using Stirling's approximation we have that $ln(K!) \geq Kln(K/e)+ln(\sqrt{2\pi K})$ and since:

$$\sum_{i=1}^{K} i \cdot ln(i) \leq \int_1^{K+1} xln(x) = x^2 \cdot (\frac{ln(x)}{2} - \frac{1}{4})\Big|_1^{K+1}$$

we obtain from Equation 3.19 that

$$\begin{aligned} &\geq \frac{1}{K^2}[(2K+1)(Kln(K/e) + ln(\sqrt{2\pi K})) - (K+1)^2 \cdot (ln(K+1) - \frac{1}{2}) + \frac{1}{2}] \\ &> ln(K) - 1.5 - \frac{K + 0.5ln(K)}{K^2} \end{aligned}$$

The claim now follows since $1.5 + \frac{K+0.5ln(K)}{K^2} < ln(5)$ for $K \geq 11$ (as assumed). □

3.7 A Lazy Version

We now turn to a more general case. We say a bin is *inactive* at a time step t if instead of permuting its content as suggested by σ^t, it disregards it and leaves it as is. We define the Markov chain $\mathfrak{M}_\alpha$ for some constant $0 < \alpha < 1$ in a similar fashion as $\mathfrak{M}$ except that in addition, the bins are independently active with probability α. It is not difficult to see that $\mathfrak{M}_\alpha$ shares the properties of ergodicity and reversibility with $\mathfrak{M}$ and therefore has $U_{\mathbb{S}_n}$ as its stationary distribution as well. Again, we would like to know $\mathfrak{M}_\alpha$'s rate of convergence.

It turns out that this modification does not have any effect on the validity of our delayed path coupling construction provided that we add to the definition of a good vertex (Definition 3.3.2) that its positions must also fall into an active bin. We do need to analyze the growth of the labeled tree in light of this modification. Also note that this modification does not relate to a vertex being isolated or not since isolation of a vertex is dependent only on the partition and has no concern with the activeness of a bin (nor the permutation). Therefore, Lemma 3.6.3 applies in this case with no modifications. On the other hand, this clearly affects the number of vertices a vertex may generate - We say a vertex $v \in V^{(t-1)}$ with $D(v) = \{p, q\}$ is *inactive* at time t if either p or q fall into an inactive bin at time t. Inactive vertices generate only a single vertex (with label $(\sigma^{(t)}(p), \sigma^{(t)}(q))$) while an active one generates as many vertices as it would under $\mathfrak{M}$. We shall show that the labeled tree growth is only slowed down by some constant factor dependent on α. We Claim the following:

Theorem 3.7.1. *For any $\epsilon > 0$, constant $\alpha > 0$, large enough n and $m \leq \frac{n}{d_0 ln(n)}$ for constant $d_0 \geq 24$*

$$\tau_{\mathfrak{M}_\alpha}(\epsilon) \leq \frac{10}{\alpha^2}(2log_{\frac{n}{10m}}(n) + 1) \cdot \frac{ln(n/\epsilon)}{ln(n) - 10}$$

We state the equivalent of Theorem 3.5.1 and proceed to detail the required modifications.

Theorem 3.7.2. *Let $i \neq j \in [n]$ and suppose $\pi_R^{(0)} = (i,j)\pi^{(0)}$. Suppose $m \leq \frac{n}{d_0 ln(n)}$ for $d_0 \geq 24$, then,*

$$\mathbb{E}[\Delta(\pi^{(T)}, \pi_R^{(T)})] \leq e \cdot n^{-1/10}$$

For $T \geq \frac{1}{\alpha^2}(2log_{\frac{n}{10m}}(n) + 1)$

The required modifications are:

1. Modification in Lemma 3.6.1: We note that in the proof of Claim 3.6.1 it is now the case that $\mathbb{E}[ln(g(v))] \geq \alpha^2 ln(\frac{n}{7.5m})$ simply since two bins are active with probability α^2 and with complement probability, $g(v) = 1$ and $ln(g(v)) = 0$. The implication of this change is that we should set D to be D/α^2 to obtain the same result as in Lemma 3.6.1.

2. Modifications in Lemma 3.6.2:

   - Consider Lemma 3.6.4. Given the s isolated vertices, it is now the case that $\mathbb{E}[X] \geq \frac{2\alpha^2 ns}{9m}$ for exactly the same reason above. Since the bins are independently active with probability α we may apply Chernoff to obtain a similar result.
   - We now seek a growth factor of at least $\frac{\alpha^2 n}{10m}$
   - We are also required that

   $$D_1 = \frac{log(\frac{m}{72ln(n)})}{log(\frac{\alpha^2 n}{10m})}$$

3. Modifications in Lemma 3.6.5: The changes are similar to that of the previous point. We require that $c_0 = \frac{\alpha^2}{342}$

4. Conditioning on $|V^{(T_0)}| > c_0 n$, each vertex is independently good with probability $\frac{\alpha}{m}$ and therefore the probability of failure in time steps $T_0 + 1, \ldots T$ is at most $n^{-\frac{\alpha c_0 n}{m}(T-T_0)}$.

Now, for large enough n and some integer constant $T - T_0 > 2\alpha^{-1}$ we can satisfy the same bounds as in the proof of Theorem 3.5.1.

Considering the fact that D, D_1 and $T - T_0$ grew each by a factor of at most α^{-2} we obtain the result in the Theorem.

Bibliography

[Ald83] D.J. Aldous. Random walks on finite groups and rapidly mixing Markov chains. *Séminaire de probabilités de Strasbourg*, 17:243–297, 1983.

[BD97] R. Bubley and M. Dyer. Path coupling: A technique for proving rapid mixing in Markov chains. In *FOCS*, pages 223–231, 1997.

[BD03] A. Beimel and D. Dolev. Buses for anonymous message delivery. *Journal of Cryptology*, 16:25–39, 2003.

[BFG+10] R. Berman, A. Fiat, M. Gomu, M. Klonowski, M. Kutyłowski, T. Levinboim, and A. Ta-Shma. Provable unlinkability against traffic analysis with low message overhead. In *Manuscript*, 2010.

[BFTS04] R. Berman, A. Fiat, and A. Ta-Shma. Provable unlinkability against traffic analysis. In *In Proc. of 8th Financial Cryptography*, pages 266–280. Spring-Verlag, 2004.

[BY] M. Blum and D. Young. Lecture Notes in Probability and Graph Theory in CS.

[Cha79] D. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. Thesis (M.S. in Computer Science), University of California, Berkeley, 1979.

[Cha81] D. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 24(2):84–88, 1981.

[Cha88] D. Chaum. The Dining Cryptographers Problem: Unconditional sender and recipient untraceability. *Journal of Cryptology*, 1(1):65–75, 1988.

[CKKL99] A. Czumaj, P. Kanarek, M. Kutyłowski, and K. Loryś. Delayed path coupling and generating random permutations via distributed stochastic processes. In *SODA*, pages 271–280, 1999.

[CKLK01] A. Czumaj, P. Kanarek, K. Lorys, and M. Kutyłowski. Switching networks for generating random permutations. *Switching Networks: Recent Advances*, 2001.

[CL05] J. Camenisch and A. Lysyanskaya. A Formal Treatment of Onion Routing. In *CRYPTO*, volume 3621 of *LNCS*, pages 169–187, 2005.

[DD06] G. Danezis and C. Diaz. A survey of anonymous communication channels. *Journal of Privacy Technology*, 20060701001, 2006.

[GKŁ04] M. Gogolewski, M. Kutyłowski, and T. Łuczak. Mobile mixing. In *Information security and cryptology (ICISC)*, volume 3506 of *LNCS*, pages 380–393, 2004.

[MR95] R. Motwani and P. Raghavan. *Randomized Algorithms*. MIT Press, 1995.

[RR98] M. K. Reiter and A. D. Rubin. Crowds: anonymity for Web transactions. *ACM Transactions on Information and System Security*, 1(1):66–92, 1998.

[RS93] C. Rackoff and D. R. Simon. Cryptographic defense against traffic analysis. In *STOC*, pages 672–681, 1993.