Problems in Discrete Probability Theory and Cryptography

By

HAMILTON SAMRAJ SANTHAKUMAR

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Mathematics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Benjamin J. Morris, Chair

Janko Gravner

Dan Romik

Committee in Charge

2022

To all my teachers.

Contents

Abstract iv

Acknowledgments vi

Chapter 1. Format Preserving Encryption Under leakage 1

1.1. Ciphers, Pseudorandom Permutations and Random Oracles 1

1.2. Introduction 3

1.3. Definition of the Cipher 5

1.4. Security of the Cipher 6

1.5. Main Technical Results 10

1.6. Randomness of the Bits Generated by Probing the Key Under Leakage 24

1.7. Comparison with the Thorp shuffle 28

1.8. The Bound on Security 34

Chapter 2. Transience of Simple Random Walks With Linear Entropy Growth 37

2.1. Introduction 37

2.2. Entropy and the Probability of Escape in $\boldsymbol{n}$ Steps 38

2.3. The Evolving Set Process 40

2.4. Relating Transience and Evolving Sets 42

2.5. Decay of $\mathbb{E}\left[\sqrt{\boldsymbol{\pi}(\boldsymbol{S}_{\boldsymbol{T}_i})}\right]$ 44

2.6. Necessity of the Uniformity of $\boldsymbol{C}$ in the Linear Entropy Growth Condition 49

Bibliography 59

Problems in Discrete Probability Theory and Cryptography

Abstract

This thesis discusses two distinct problems, the first of which is concerning format preserving encryption in cryptography and the second is concerning transience of simple random walks on infinite graphs.

In Chapter 1, we discuss format preserving encryption (FPE) under the presence of an adversary who can leak parts of the secret key. The main aim of FPE is to encrypt a plaintext into a ciphertext of the same format. For example, under FPE, an encrypted credit card number will still look like a credit card number. One way to achieve this is by generating a uniformly random permutation on the space of all plaintexts and then applying this permutation on the plaintexts to get ciphertexts. Storing such a permutation is infeasible. For example in the case of credit cards, it would take $61,391$ terabytes to store such a permutation. So instead, in cryptography what we usually look for is a random permutation which is an easy to compute function of a typically short random string called the key. Such a random permutation is never going to be equal to a uniformly random permutation in distribution as long as the key size is small. All one needs is that it is practically very hard to distinguish between the two. Such a construction only works as long as the key is secret, because knowing the key is same as knowing the permutation. In [**7**], Bellare, Kane and Rogaway provide an encryption scheme that is secure even in the presence of an adversary who has partial knowledge of the key. They thwart such an adversary by making the key large and putting an upper limit on the amount of information about the key that the adversary can steal. The rationale behind this is that for a threat residing in one's network, it is hard for it to transfer huge amounts of data to an external location without being detected. Their encryption scheme isn't format preserving. In this chapter, we discuss format preserving encryption in the same setting as [**7**]. In particular we provide a format preserving encryption scheme and prove that it is secure under an appropriately modified notion of security.

In chapter 2 we explore the connection between the entropy growth of a simple random walk on a connected infinite graph with bounded degree and its transience. In particular, using the technique of evolving sets we show that for a simple random walk starting at any vertex of an

infinite connected graph with bounded degree, if the entropy of its n^{th} step grows at least linearly in n, with the constant of linearity being independent of the starting position, then the random walk is transient. For irreducible transitive Markov chains, it is already known that linear entropy growth implies transience. So, we end up enlarging the class of chains for which this result holds. We also give an example to show that this uniformity of the constant of linearity is an essential condition. That is, if there is no constant of linearity which works for every starting position, then the random walk is not necessarily transient.

Acknowledgments

First and foremost, I would like to thank my advisor Prof. Ben Morris for his unwavering support and expert guidance. I've learned a lot from him over the years, both about mathematics and life in general. If I emulate even a portion of his impressive intuition for probability theory, I'm sure I'll be able to find success in any future mathematical endeavors. I'm thankful for his immense patience and kindness. Without him, this work would not have been possible.

I would like to thank Prof. Janko Gravner and Prof. Dan Romik for agreeing to be on both my dissertation committee and my qualifying exam committee. I would like to thank fellow graduate student and collaborator Hans Oberschelp for collaborating with me on our cryptography project.

I would like to thank two of my closest friends, Ojesh Koul and Robert Scherer, for their constant support and encouragement. They have taught me a lot about life and their kind words have helped me greatly during tough times. I would like to thank my sister, Jelin who has been a source of inspiration for me through her display of tenacity and hard work.

I would like to thank the first floor administration, especially Tina Denena, Sarah Driver and Victoria Whistler, who have been there for me whenever I got lost in the administrative aspects of graduate school. Last but not the least, I'm thankful to all the Professors who have taught me mathematics through the years.

CHAPTER 1

Format Preserving Encryption Under leakage

1.1. Ciphers, Pseudorandom Permutations and Random Oracles

In this section, we provide some preliminaries from the subject of cryptography, namely we introduce the concept of pseudorandom permutations and the random oracle model. A cipher is a function $E : \mathscr{K} \times \mathscr{M} \to \mathscr{M}$ such that $E(x, \cdot)$ is a permutation on $\mathscr{M}$ for every x. Elements in $\mathscr{K}$ are called keys. An element in $\mathscr{M}$ is either called plaintext or a ciphertext based on whether it is in the domain or co-domain respectively. Plaintexts will be interchangeably referred to as messages throughout this chapter. Often times, $\mathscr{K}$ and $\mathscr{M}$ are sets of strings of fixed lengths. The aim of format preserving cryptography is to send a message $m \in \mathscr{M}$ to the receiver by changing it into $m' \in \mathscr{M}$ while an adversary is watching and trying to obtain m from m'. The function $m \to m'$ must be invertible, otherwise there is no way for the receiver to read the original message. So the best possible way to do this is for the sender and receiver to meetup once and generate an instance of a random permutation π on $\mathscr{M}$. Then in the future one can transmit $\pi(m)$ to the sender. The permutation chosen is kept as a secret and unknown to the adversary. Since the permutation is completely random, the adversary cannot do anything better than guessing m from m'. However, as we discussed in the abstract, this is infeasible due to the costs involved in storing such a permutation. Hence, instead of generating a random permutation, the sender and receiver generate a shared key $K \in \mathscr{K}$ and use the permutation $E(K, \cdot)$ to transfer messages among themselves. Naturally, since we are trying to emulate uniformly random permutations, we want $E(K, \cdot)$ to be close to one in some sense. If $E(K, \cdot)$ is close to a uniformly random permutation in some fixed measure of closeness, we call it a *pseudorandom permutation*. The measure used to describe this closeness of random permutations depends on applications. We will now describe two such notions, namely the CCA (chosen ciphertext attack) and CPA (chosen plaintext attack) security for pseudorandom permutations. Let E be a cipher and let K be a uniformly chosen key.

Consider the following two worlds and place the adversary in each of these one at a time.

World 0: Choose a uniformly random permutation $\pi : \mathscr{M} \to \mathscr{M}$ and set $g = \pi$. Give the adversary a black box access to g and g^{-1}. What this means is that the adversary doesn't get any description of the functions but rather gets two black boxes. He can enter an $m \in \mathscr{M}$ into either of these black boxes and get the corresponding output. He is only allowed to do this operation q times. These operations are called queries.

World 1: In this world, set $g = E(K, \cdot)$ and just as in world 0, give the adversary black box access to both g and g^{-1}. Again, he is allowed to make a total of q queries.

In each of these worlds, we ask the adversary to make an educated guess about the world he is in. Let $\mathscr{A}(i)$ be the answer he gives while in world i. The CCA advantage of an adversary $\mathscr{A}$ is defined to be

$$\mathrm{Adv}^{\mathrm{CCA}}(\mathscr{A}) = \mathbb{P}\big(\mathscr{A}(1) = 1\big) - \mathbb{P}\big(\mathscr{A}(0) = 1\big).$$

The maximum CCA advantage is defined to be

$$\mathrm{MaxAdv}_{\mathrm{q}}^{\mathrm{CCA}} = \max_{\mathscr{A}} \big(\mathrm{Adv}^{\mathrm{CCA}}(\mathscr{A})\big),$$

where the max is taken over all adversaries who are allowed to make at most q queries. We say that a pseudorandom permutation is secure against CCA if the above maximum advantage is small. If in worlds 0 and 1, the adversary is only allowed black box access to g, then one gets the notion of security against CPA. Security against CCA and CPA are standard in the literature. See [**6**] for example. We will be using a notion of security that is a modification of CCA/CPA. Refer Section 1.4 for the definitions that we use. The modification was suggested to us by Phillip Rogaway (personal communication, 2019). In the remaining part of this section, we will go over the random oracle model of cryptography.

All the proofs in this chapter are in a model called the *random oracle model*, which was first used in cryptographic proofs by Bellare and Rogaway in [**8**]. Given a finite domain A and co-domain B, the random oracle model assumes the existence of a function f which produces a uniformly random element from B when applied to an element $x \in A$, such that this random element is independent of the values of the function at points other than x. When this function is applied on x once again, it produces the same output as before. Mathematically, what it means is that for

$A = \{x_1, \ldots, x_u\}$, the sequence $f(x_1), \ldots, f(x_u)$ is a single instance of a sequence of i.i.d. random elements that are uniform over B. In the literature, f is typically defined as a single instance of a random function that is uniform over the collection of functions with domain A and co-domain B. These two depictions of the function f are equivalent. f is called a random oracle and $f(x)$ is called a random oracle call. Typically when one talks about random oracles, the domain and co-domain are suppressed. So it is a common practice to say for example, "apply the random oracle on the string 0110 to get a uniformly random string of length 10". We will be using this language throughout this chapter. Random oracles are highly idealized representations of cryptographic hash functions. Without the random oracle model, proofs in cryptography become incredibly hard. We will not be questioning the validity of the random oracle model since it falls outside the realms of mathematics.

It is important to note that all parties involved including the adversary have access to the random oracles. Proofs in the random oracle model use the idea that until a random oracle call $f(x)$ is made by a party, $f(x)$ is completely random in the eyes of the party. Once the call is made and its value is revealed, $f(x)$ becomes deterministic. However, revealing the value of $f(x)$ doesn't give any information about $f(y)$ for $y \neq x$. This is because a random oracle produces outputs that are independent of each other. Practically, the number of random oracle calls that can be made by an adversary is limited. So, in the random oracle model one assumes that the adversary can make at most r random oracle calls and computes security bounds that depend on r.

1.2. Introduction

The main aim of this chapter is to solve the problem of format preserving encryption in the bounded retrieval model, by constructing a pseudorandom permutation and providing concrete security bounds in the random oracle model. The bounded retrieval model was introduced to study cryptographic protocols that remain secure in the presence of an adversary that can transmit or leak private information including the key from the host's computer to a remote home base. One example of such an adversary is an APT (Advanced Persistent Threat), which is a malware that stays undetected in the host's network and tries to ex-filtrate the secret keys used by the host. The premise of the bounded retrieval model is that such an adversary cannot move a large amount of

data to a remote base without being detected or that it can only communicate with the remote base through a very narrow channel. That is, the model assumes an upper bound on the amount of data that an adversary can leak. For a list of works that are set in the bounded retrieval model, refer to [**1**, **2**, **7**, **9**, **12**, **15**]. In [**7**] Bellare, Kane and Rogaway introduce an efficient symmetric encryption scheme in this model and give concrete security bounds for it. They assume that the secret key is very large and model the leaked data as a function that takes the secret key as the input and outputs a smaller string. The length of this string is a parameter on which the security bounds depend. Their algorithm uses a random seed R along with the big key to generate a key of conventional length that is indistinguishable from a random string of the same length even when the function used to model the leaked data depends on calls to the random oracle that the algorithm uses. It then uses this newly generated key and any of the conventionally available symmetric encryption schemes, say an AES mode of operation, to create a ciphertext C. Finally it outputs (R, C).

The above scheme is not format preserving since the final ciphertext (R, C) is longer than the original message M. A question posed by Phillip Rogaway (personal communication, 2019) is whether a secure format preserving encryption scheme exists in the bounded retrieval model. Another way to pose this question is as follows : If the adversary is allowed to leak data, is it possible to construct a pseudorandom permutation that is secure under some notion of security, say the CCA notion of security? The aim of this chapter is to answer this question. Unfortunately it is not possible to come up with a pseudorandom permutation that is secure under the strong notion of CCA security. This is because in the CCA model, before trying to distinguish between a random permutation and the pseudorandom permutation, the adversary can choose to look at a sequence of plaintext-ciphertext pairs that he chooses. If a leakage of data is allowed, the adversary can simply leak a single plaintext-ciphertext pair and use it to gain a very high CCA advantage. Hence we weaken the notion of security by requiring that the adversary can only look at a sequence of plaintext-ciphertext pairs where the plaintexts are uniformly random and distinct. We then ask him to distinguish between a truly random permutation and the pseudorandom permutation. In this chapter we give a pseudorandom permutation in the bounded retrieval model and prove that it is secure in the weak sense that is discussed above. The precise defintion of security in our setup can be found in Section 1.4.

Just as in [**7**] we utilize a big key. The main idea in this chapter is that if one fixes the string of leaked bits, the key is a uniform sample from the preimage of the leaked string. If the length of the leaked string is small, then on average the preimage is very large. What this means is that even when the leakage is known, with a high probability the total entropy of the key is high. This implies that the sum of entropies of each bit in our key is large. This intuitively means that if some of the bits in the key are not very random when the leakage is known, the other bits must somewhat resemble unbiased random bits. So, if one uses a random oracle to look at various positions of the key and take an XOR, it is likely that the resulting bit is close to an unbiased random bit. This idea of probing the key is similar to the one used in [**7**]. The content of Sections 1.5 and 1.6, which form the heart of this chapter, is to show that bits generated by probing the key are close to i.i.d. unbiased random bits. To construct a pseudorandom permutation using these bits, we use a particular card shuffling scheme called the Thorp shuffle, just as in [**25**]. This construction is given in the next section.

1.3. Definition of the Cipher

Let $\mathscr{M} = \{0,1\}^m$ be the set of messages and $\mathscr{K} = \{0,1\}^k$ the set of keys. For a given message $M \in \mathscr{M}$ and key $K \in \mathscr{K}$, set $\mathrm{Al}_0(M) = M$ and define $\mathrm{Al}_t(M)$ inductively as follows. If $\mathrm{Al}_{t-1}(M) = b_1 b_2 \dots b_m$, then apply the random oracle on $(b_2, b_3, \dots, b_m, t)$ to obtain $\big((P_1, P_2, \dots, P_n), \mathscr{S}\big)$, where $P_1, \dots, P_n$ are distinct uniform samples from $\{1, \dots, k\}$ and $\mathscr{S}$ is a uniformly chosen random subset of $\{1, \dots, n\}$ that is independent of $(P_1, \dots, P_n)$. Let $c = \bigoplus_{i \in \mathscr{S}} K[P_i]$, where $K[P_i]$ denotes the bit at the P_i^{th} position of the key K. Set $\mathrm{Al}_t(M) = b_2 b_3 \dots b_m (c \oplus b_1)$. Finally, we define the cipher $E : \mathscr{K} \times \mathscr{M} \to \mathscr{M}$ to be $E(K, M) = \mathrm{Al}_T(M)$ where T is some fixed number. We will call $P_1, \dots, P_n$ *probes* and $\mathscr{S}$ *sub probe indices*.

To put it in words, we generate random bits by probing the key using a random oracle and then taking the XOR of a subset of bits in these probe positions. We then use these random bits to do T steps of the Thorp shuffle on the space of messages. Thorp shuffle has been already used by Morris, Rogaway and Stegers in [**25**] to build a pseudorandom permutation on the space of binary strings of desired length from a pseudorandom function. This works to our advantage because now

it is enough to show that it is hard to distinguish between our pseudorandom permutation and the pseudorandom permutation from [**25**]. This is the content of Section 1.7.

1.4. Security of the Cipher

In this section we introduce a notion of security for pseudorandom permutations under the assumption that there is a leakage of data. We assume that the adversary can leak l bits of data and just as in [**7**], use a function $\Phi : \mathscr{K} \to \{0,1\}^l$ to model this. Henceforth we will refer to this function as the *leakage function*. The adversary has the power to choose this function and this function can depend on calls to the random oracle. For a key K, we will use $L = \Phi(K)$ to denote the output one gets by applying the leakage function to it. We will call this the *leakage*. We allow the adversary to make r random oracle calls and decide on a leakage function Φ. After the adversary has chosen a leakage function, consider the following two worlds.

World 1: In this world, we first choose distinct uniformly random messages $M_1, \ldots, M_q \in \mathscr{M}$. Then, for a uniformly random key $K \in \mathscr{K}$, we set $C_i = E(K, M_i)$ where E is the cipher we defined in Section 1.3. We give the adversary access to the leakage L, the input-output pairs $(M_1, C_1), \ldots, (M_q, C_q)$ and the random oracle calls that were used by the algorithm to compute the $E(K, M_i)'s$.

World 0: In this world, again we choose distinct uniformly random messages $M_1, \ldots, M_q \in \mathscr{M}$. We once again choose a random key $K \in \mathscr{K}$ and compute L and all the random oracle calls necessary to evaluate the $E(K, M_i)'s$, just like world 1. However, instead of setting $C_i's$ to be the outputs of the cipher, we do the following. We choose a uniformly random permutation $\pi : \mathscr{M} \to \mathscr{M}$ and set $C_i = \pi(M_i)$. Just as in world 1, the adversary is provided access to the input-output pairs for the q messages, the leakage L and the random oracle calls.

We now place him in these two worlds one at a time without telling him which world he is in. In each of these cases we ask the adversary to guess which world he is in. Let $\mathscr{A}(0)$ and $\mathscr{A}(1)$ denote the answers he gives in world 0 and world 1 respectively. Then, we define the advantage of an adversary as

$$\mathbf{Adv}(\mathscr{A}) = \mathbb{P}_1\big(\mathscr{A}(1) = 1\big) - \mathbb{P}_0\big(\mathscr{A}(0) = 1\big), \tag{1.1}$$

where $\mathbb{P}_i$ is the probability measure in world i. Define the maximum advantage

$$\mathbf{MaxAdv_{r,q}} = \max_{\mathscr{A}} \Big(\mathbf{Adv}(\mathscr{A})\Big), \tag{1.2}$$

where the maximum is taken over all adversaries satisfying the above mentioned conditions. Note that in the above setup if we allow the messages to be chosen by the adversary instead of them being random, we get the notion of security against chosen plain text attack (CPA) under leakage. Security against CPA is weaker than security against CCA (chosen ciphertext attack). Unfortunately, if a leakage is allowed, it is not possible to design a cipher that is secure in the CPA framework. This is because of the adversary who does the following: Let $q = 1$. Assume that the message length m is less than l. For each key K, the adversary includes the ciphertext $E(K, M_1)$ into the leakage, for a fixed message M_1. Then, the adversary answers as follows. If $C_1 = E(K, M_1)$ then the adversary guesses that he is in world 1. Else, the guess is world 0. In this case, $\mathbb{P}_1(\mathscr{A}(1) = 1) = 1$ and $\mathbb{P}_0(\mathscr{A}(0) = 1) = 1/2^m$. Hence this adversary has a very high advantage. There is an analogue to this strategy even in our setup, which we will call the naive strategy. We will discuss it after stating the main theorem. The following notation is useful when we make a comparison between our bound and the naive strategy. Let $M_1, \ldots, M_j$ be distinct uniform samples from $\mathscr{M} = \{0,1\}^m$. Then for any subset $\mathscr{M}' \subseteq \mathscr{M}$ with $|\mathscr{M}'| = h$, define

$$B(h, j, m) = \mathbb{P}\big(M_i \in \mathscr{M}' \text{ for some } 1 \leq i \leq j\big). \tag{1.3}$$

Note that $B(h, j, m)$ is increasing in h and that one can bound $B(h, j, m)$ using an union bound in the following way.

$$B(h, j, m) \leq \sum_{i=1}^{j} \mathbb{P}\big(M_i \in \mathscr{M}'\big) = j\frac{|\mathscr{M}'|}{2^m} = \frac{hj}{2^m}.$$

The main result of this chapter is the following theorem which is a bound on the maximum advantage.

THEOREM 1.4.1. *Let* $\mathbf{MaxAdv_{r,q}}$ *be as above. Let* q *be the number of messages available to the adversary in world 0 or 1 and let* l, k, T, n, m *be as in Section 1.3. Let* s *be a whole number such that* $T = s(2m-1)$ *and let* r *be the maximum number of random oracle calls that the adversary*

can make while deciding on a leakage function. Assume that $l + qT + n/(2\ln 2) \leq k - n$*, then*

$$\mathbf{MaxAdv}_{\mathbf{r,q}} \leq 2qT \cdot 2^{-n\cdot\alpha(k,l+qT-1,n)} + \frac{q}{s+1}\left(\frac{4mq}{2^m}\right)^s + B(2r,q,m), \tag{1.4}$$

where

$$\alpha(k,a,n) = \frac{1}{8\ln 2}\left(1 - \frac{a+1+n/(2\ln 2)}{k-n}\right)^2 \tag{1.5}$$

and the function B *is as defined in* (1.3)*.*

To make sense of the above result, let's consider the following strategy which we will call the *naive strategy*. For given l, k set $u = \lfloor l/m \rfloor$. Then for any u distinct messages compute the corresponding ciphertexts. Call the collection of message used $\mathscr{M}'$. Now set the leakage to be a concatenation of the u ciphertexts obtained, followed by any $l - m\lfloor l/m \rfloor$ bits. Next, when placed in either world 0 or world 1, check whether any of the q random messages provided is from the collection $\mathscr{M}'$. Say M_i is from this collection. Then the ciphertext for M_i for is known. Call this ciphertext C_i. Answer world 1 if the output corresponding to message M_i is C_i, else answer 0. If none of the q random messages is from the collection $\mathscr{M}'$, then answer based on an independent unbiased coin toss. Let $\mathbf{Adv}_{\mathbf{naive}}$ denote the advantage of this strategy. Then,

$$\mathbf{Adv}_{\mathbf{naive}} = B(\lfloor l/m \rfloor, q, m)(1 - 1/2^m).$$

The maximum number of random oracle calls made in this strategy is $\lfloor l/m \rfloor T$, one for each message and time step. So, it makes sense to compare the naive advantage to the bound we get when $r = \lfloor l/m \rfloor T$. Let $\mathscr{M}_1, \ldots, \mathscr{M}_h$ be disjoint subsets of $\mathscr{M}$. Set $a_i = |\mathscr{M}_i|$. Then,

$$B\left(\sum_{v=1}^{h} a_v, j, m\right) = \mathbb{P}\left(M_i \in \bigcup_{v=1}^{h} \mathscr{M}_v \text{ for some } i\right) = \mathbb{P}\left(\bigcup_{v=1}^{h} \left\{M_i \in \mathscr{M}_v \text{ for some } i\right\}\right)$$
$$\leq \sum_{v=1}^{h} \mathbb{P}\left(M_i \in \mathscr{M}_v \text{ for some } i\right) = \sum_{v=1}^{h} B(a_v, j, m).$$

This gives us

$$B(2T\lfloor l/m \rfloor, q, m) \leq 2T \cdot B(\lfloor l/m \rfloor, q, m).$$

So, if we set $r = \lfloor l/m \rfloor T$, we have

$$\mathbf{MaxAdv_{r,q}} \leq 2qT \cdot 2^{-n \cdot \alpha(k, l+qT-1, n)} + \frac{q}{s+1}\left(\frac{4mq}{2^m}\right)^s + 2T \cdot \mathbf{Adv_{naive}}.$$

So, at least in the special case of $r = \lfloor l/m \rfloor$, we can conclude that the advantage of the best strategy is at most $2T$ times the advantage of the naive strategy plus an additional error term. Let's plot this error term for some concrete values. Fix $k = 2^{43}$, which means the key is 1 Terabyte long. Fix $l = k/8 = 2^{40}$, i.e., 12.5% or about 125 gigabytes of the key is allowed to be leaked. Fix the message length to be $m = 64$ and the number of probes to be $n = 500$. Fix $T = s(2m-1)$ with $s = 2$, i.e., $T = 254$. Let

$$f(q) = 2qT \cdot 2^{-n \cdot \alpha(k, l+qT-1, n)} + \frac{q}{s+1}\left(\frac{4mq}{2^m}\right)^s$$

and let

$$\Gamma(q) = \min(f(q), 1).$$

Figure 1.1 shows a plot with $\log_2(q)$ on the x-axis and $-\log_2(\Gamma(q))$ on the y-axis.

General outline for the proof of the main theorem: The proof of the main theorem is broken down into three steps. The first step is to show that if the random oracle call necessary to compute the probes for a certain bit is not made by the adversary, then on average, the bit obtained is random even when one conditions on the leakage. To show this we note that given a leakage, $L = \Phi(K)$, the distribution of keys is uniform over $\Phi^{-1}(L)$. We know that the size of this set is typically 2^{k-l}, or more precisely that $\mathbb{E}\big(1/|\Phi^{-1}(L)|\big) = 2^{-(k-l)}$. So, we prove a bound on the randomness of the bit assuming that the size of $\Phi^{-1}(L)$ is fixed. This easily translates to a bound for the case when the size of $\Phi^{-1}(L)$ is not fixed, by an application of Markov's inequality. The second step is to use the randomness of bits obtained by probing to show that the result of applying our algorithm is close in total variation distance to the result obtained by applying the Throp shuffle. Pseudorandom permutations that are close in total variation distance are hard to distinguish from each other. This step also assumes that the random oracle calls required for applying our algorithm to the inputs in question were not used to decide on a leakage function.

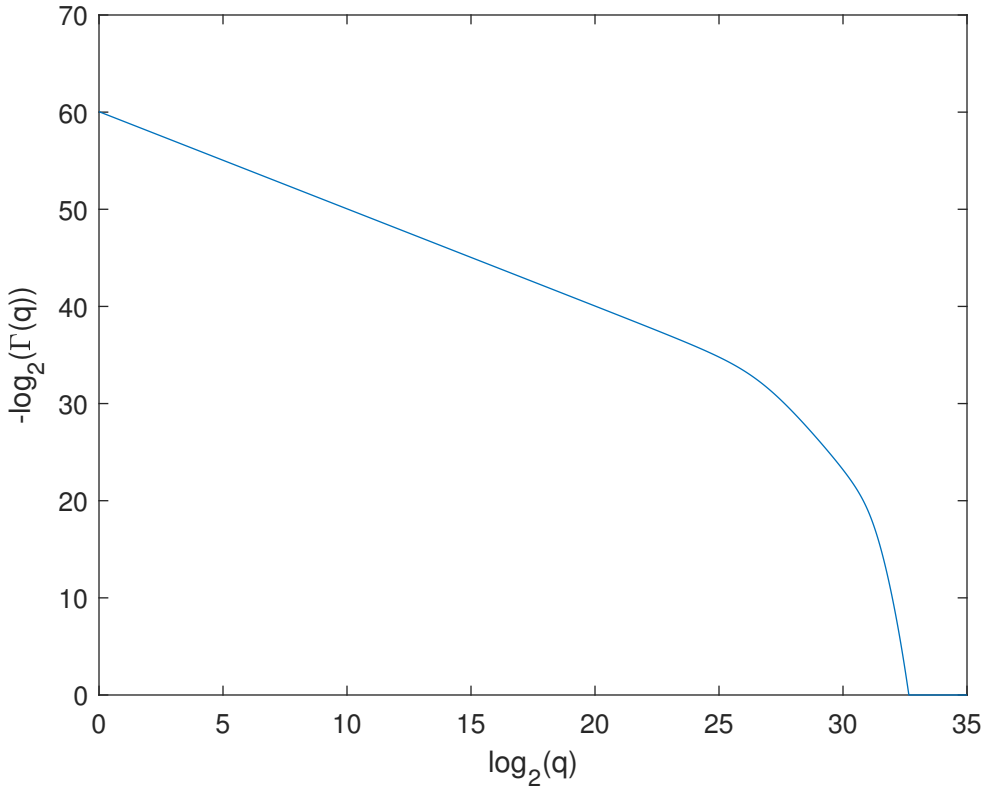

FIGURE 1.1. A plot of $\log_2(q)$ vs $-\log_2\big(\Gamma(q)\big)$

The third and final step is to extend it to the case where the adversary makes any random oracle calls he wants, by simply assuming that if the adversary uses the random oracle on $(b_2, \ldots, b_m, t)$ where $C = b_1 \ldots b_m$ is the ciphertext at round/time t corresponding to a message M, and $a = 0$ or 1, then the message M has been compromised. By compromised we mean that the cipher text of this message is no longer random, i.e., it is completely known.

1.5. Main Technical Results

Let $P_1, P_2, \ldots, P_n$ and $\mathcal{S}$ be the random probes and sub probes indices as in Section 1.3. Throughout this section, by an abuse of notation, we will use K to denote an arbitrary random string of length k, instead of the key. Most of the time we will assume that $\mathbb{P}(K = x) \leq 2^{-(k-d)}$ for all x. This is weaker than the condition that K is sampled from a set of size 2^{k-d}. We will also assume that the probes and sub probe indices are independent of this K. Let $c = \bigoplus_{i \in \mathcal{S}} K[P_i]$. We

begin by relating the randomness of c with the distribution of $(K[P_1], K[P_2], \ldots, K[P_m])$ through the following lemma.

LEMMA 1.5.1. *Let $Y = (Y_1, Y_2, \ldots, Y_n) \in \{0,1\}^n$ be a random n-bit string. For $S \subseteq \{1, \ldots, n\}$, set $f_S(Y) := (-1)^{\oplus_{i \in S} Y_i}$, with the convention that $f_\emptyset \equiv 1$. Also let*

$$\mathscr{E}(S) = \mathbb{E}[f_S(Y)].$$

Then for a uniformly chosen random subset $\mathcal{S} \subseteq \{1, \ldots, n\}$, we have

$$\mathbb{E}\big[\mathscr{E}(\mathcal{S})^2\big] = \sum_{y \in \{0,1\}^n} \mathbb{P}\big(Y = y\big)^2. \tag{1.6}$$

PROOF. Let $\Omega = \{0,1\}^n$. Note that $\mathbb{R}^\Omega$, the space of real valued functions on Ω, forms a vector space of dimension $|\Omega|$ over $\mathbb{R}$. Define the following inner product on $\mathbb{R}^\Omega$.

$$\begin{aligned}\langle f, g\rangle &= \frac{1}{2^n} \sum_{x \in \Omega} f(x) g(x), \text{ for } f, g \in \mathbb{R}^\Omega \\ &= \mathbb{E}[f(Z) g(Z)],\end{aligned}$$

where $Z = (Z_1, \ldots, Z_n)$ and $Z_1, Z_2, \ldots, Z_n$ are i.i.d Bernoulli(1/2) random variables. Observe that when $S \neq S'$,

$$\begin{aligned}\langle f_S, f_{S'}\rangle = \mathbb{E}[f_S(Z) f_{S'}(Z)] &= \mathbb{E}\left[\prod_{i \in S \cap S'} (-1)^{2Z_i} \prod_{j \in S \triangle S'} (-1)^{Z_j}\right] \\ &= \prod_{i \in S \cap S'} \mathbb{E}\Big[(-1)^{2Z_i}\Big] \prod_{j \in S \triangle S'} \mathbb{E}\Big[(-1)^{Z_j}\Big] = 0,\end{aligned}$$

since $\mathbb{E}[(-1)^{Z_i}] = 0$ and $S \triangle S'$ is non-empty when $S \neq S'$. Also observe that

$$\langle f_S, f_S\rangle = \mathbb{E}\Big[(-1)^{2(\oplus_{i \in S} Z_i)}\Big] = \mathbb{E}[1] = 1.$$

Therefore, $\{f_S\}_{S \in 2^{[n]}}$ forms an orthonormal basis for 2^Ω. Next, let $U(y) = 1/2^n$ and $P(y) = \mathbb{P}\big(Y = y\big)$ for $y \in \Omega$. Then, $P, U, P/U \in \mathbb{R}^\Omega$. Now note that

$$\langle P/U, f_S\rangle = \frac{1}{2^n} \sum_{x \in \Omega} 2^n P(x) f_S(x) = \sum_{x \in \Omega} P(x) f_S(x) = \mathbb{E}[f_S(Y)] = \mathscr{E}(S)$$

$$\implies \frac{1}{2^n}\langle P/U, f_S\rangle^2 = \frac{1}{2^n}\mathscr{E}(S)^2.$$

Summing the above equation over all subsets $S \subseteq \{1,\dots,n\}$ and using the fact that $f'_S s$ form an orthonormal basis, we get

$$\frac{1}{2^n}\langle P/U, P/U\rangle = \sum_{S\subseteq\{1,\dots,n\}} \frac{1}{2^n}\langle P/U, f_S\rangle^2 = \sum_{S\subseteq\{1,\dots,n\}} \frac{1}{2^n}\mathscr{E}(S)^2 = \mathbb{E}\big[\mathscr{E}(\mathcal{S})^2\big].$$

The left hand side of the above equation simplifies to $\sum_{y\in\Omega} P(y)^2$ and hence the proof is complete.

□

Notation. Let X, Y be two random variables. Then let $\mathscr{L}(X)$ and $\mathscr{L}(X|Y)$ denote the law of X and the law of X given Y respectively.

Definition (Total variation distance)**.** Let μ and ν be two probability distributions supported on a finite set Ω. Then the total variation distance between them is denoted by,

$$\|\mu - \nu\|_{TV} = \frac{1}{2}\sum_{x\in\Omega} |\mu(x) - \nu(x)|. \tag{1.7}$$

Corollary 1.5.1. *Let K be a random string. Let $\mathscr{P}^n = (P_1,\dots,P_n)$ be probes and let $\mathcal{S}$ be random sub probe indices, as defined in Section 1.3. Assume that $K, \mathscr{P}^n$ and $\mathcal{S}$ are independent of each other. Set $Y = (Y_1,\dots,Y_n)$ where $Y_i = K[P_i]$. Then for $c = \oplus_{i\in\mathcal{S}} K[P_i]$ and $c' \sim$ Bernoulli(1/2)*

$$\mathbb{E}\big[\|\mathscr{L}(c\ |\mathscr{P}^n, \mathcal{S}) - \mathscr{L}(c')\|_{TV}\big] \le \frac{1}{2}\mathbb{E}\left[\sqrt{\sum_{y\in\{0,1\}^n} \mathbb{P}\left(Y = y \mid \mathscr{P}^n\right)^2}\right].$$

Proof. Condition on the event $\mathscr{P}^n = \mathpzc{p}$. Let $f_S(\cdot)$ be as in Lemma 1.5.1. We will apply Lemma 1.5.1 to the probability distributions obtained by conditioning. So it is useful to define $\widetilde{\mathbb{E}}(\cdot) = \mathbb{E}[\cdot \mid \mathscr{P}^n = \mathpzc{p}]$. Begin by observing that

$$\begin{aligned}\mathbb{E}\big[f_{\mathcal{S}}(Y) \mid \mathscr{P}^n = \mathpzc{p}, \mathcal{S} = S\big] &= \sum_{y\in\{0,1\}^n} f_S(y)\mathbb{P}\big(Y = y \mid \mathscr{P}^n = \mathpzc{p}, \mathcal{S} = S\big)\\ &= \sum_{y\in\{0,1\}^n} f_S(y)\mathbb{P}\big(Y = y \mid \mathscr{P}^n = \mathpzc{p}\big)\end{aligned}$$

since $Y = (K[P_1], \ldots, K[P_n])$ and $\mathscr{P}^n$ are independent of $\mathcal{S}$. So we get

$$\mathbb{E}\big[f_{\mathcal{S}}(Y) \mid \mathscr{P}^n = \mathscr{p}, \mathcal{S} = S\big] = \widetilde{\mathbb{E}}[f_S(Y)]. \tag{1.8}$$

Define $\mathscr{E}(S) = \widetilde{\mathbb{E}}[f_S(Y)]$. Note that $\mathcal{S}$ is uniformly distributed even after conditioning, due to independence. Hence we can apply Lemma 1.5.1 to the conditioned random variables Y and $\mathcal{S}$ to get

$$\widetilde{\mathbb{E}}\big[\mathscr{E}(\mathcal{S})^2\big] = \sum_{y \in \{0,1\}^n} \mathbb{P}\big(Y = y \mid \mathscr{P}^n = \mathscr{p}\big)^2.$$

Applying Jensen's inequality, we get

$$\widetilde{\mathbb{E}}\big[|\mathscr{E}(\mathcal{S})|\big] \le \sqrt{\widetilde{\mathbb{E}}[\mathscr{E}(\mathcal{S})^2]} = \sqrt{\sum_{y \in \{0,1\}^n} \mathbb{P}\big(Y = y \mid \mathscr{P}^n = \mathscr{p}\big)^2}. \tag{1.9}$$

For Bernoulli distributions, one can check that $\|\text{Bernoulli}(q') - \text{Bernoulli}(1/2)\|_{TV} = |q' - 1/2|$. So for $q = \mathbb{P}\big(c = 1 \mid \mathscr{P}^n = \mathscr{p}, \mathcal{S} = S\big)$, we have

$$\Big\|\mathscr{L}\big(c \mid \mathscr{P}^n = \mathscr{p}, \mathcal{S} = S\big) - \mathscr{L}(c')\Big\|_{TV} = |1/2 - q|.$$

Also note that

$$|1/2 - q| = \Big|(1/2 - 1)q + (1/2 - 0)(1 - q)\Big| = \Big|\mathbb{E}\big[1/2 - c \mid \mathscr{P}^n = \mathscr{p}, \mathcal{S} = S\big]\Big|.$$

Finally observe that $c = \big[1 - f_{\mathcal{S}}(Y)\big]/2$ since $\oplus_{i \in \mathcal{S}} K[p_i] = 0$ implies $f_{\mathcal{S}}(Y) = 1$ and $\oplus_{i \in \mathcal{S}} K[p_i] = 1$ implies $f_{\mathcal{S}}(Y) = -1$. So the above two equations give us

$$\begin{aligned}
\Big\|\mathscr{L}\big(c \mid \mathscr{P}^n = \mathscr{p}, \mathcal{S} = S\big) - \mathscr{L}(c')\Big\|_{TV} &= \Big|\mathbb{E}\big[1/2 - c \mid \mathscr{P}^n = \mathscr{p}, \mathcal{S} = S\big]\Big| \\
&= \frac{1}{2}\Big|\mathbb{E}\big[f_{\mathcal{S}}(Y) \mid \mathscr{P}^n = \mathscr{p}, \mathcal{S} = S\big]\Big| \\
&= \frac{1}{2}\Big|\widetilde{\mathbb{E}}\big[f_S(Y)\big]\Big| \text{ from (1.8)} \\
&= \frac{1}{2}\big|\mathscr{E}(S)\big|.
\end{aligned}$$

Using the above equation and inequality (1.9), we get

$$\mathbb{E}\left[\left\|\mathscr{L}\big(c \mid \mathscr{P}^n, \mathcal{S}\big) - \mathscr{L}(c')\right\|_{TV} \;\middle|\; \mathscr{P}^n = \boldsymbol{p}\right] \leq \frac{1}{2}\sqrt{\sum_{y\in\{0,1\}^n} \mathbb{P}\big(Y = y \mid \mathscr{P}^n = \boldsymbol{p}\big)^2}.$$

This is true for every $\boldsymbol{p}$. So we can take an expectation with respect to $\mathscr{P}^n$ to finish the proof of the corollary. □

This shows that the average randomness of the bit c depends on the average l^2-norm of the vector of probabilities defined by $(K[P_1], \ldots, K[P_n])$, where the average is taken over the probes. Our aim is to bound this quantity assuming that they key K is sampled from a large enough subset $\mathscr{K}' \subseteq \mathscr{K}$. The theorem below gives us a bound under a weaker assumption.

THEOREM 1.5.1. *Assume that a random string $K \in \{0,1\}^k$ satisfies*

$$\mathbb{P}(K = y) \leq \frac{1}{2^{k-d}} \text{ for all } y \text{ and some } d \geq 0. \tag{1.10}$$

Assume that the probes $P_1, \ldots, P_n$ are independent of K and that $d + 1 + n/(4\ln 2) \leq k - n$. Then for $Y = (K[P_1], K[P_2], \ldots, K[P_n])$,

$$\mathbb{E}\left[\sum_{y\in\{0,1\}^n} \mathbb{P}\left(Y = y \mid P_1, P_2, \ldots, P_n\right)^2\right] \leq 2^{-2n\cdot\beta(k,d,n)}, \text{ where} \tag{1.11}$$

$$\beta(k,d,n) = \frac{1}{8\ln 2}\left(1 - \frac{d + 1 + n/(4\ln 2)}{k-n}\right)^2.$$

We need a series of lemmas before we can prove this. So we postpone the proof to the end of this section. Our first step in this regard is using Jensen's inequality to convert the hypothesis of the above theorem into a lower bound on the entropy of bits in K. This is the content of the lemma given below.

LEMMA 1.5.2. *Assume that a random string $K \in \{0,1\}^k$ satisfies*

$$\mathbb{P}(K = x) \leq \frac{1}{2^{k-d}} \text{ for all } x \text{ and some } d \geq 0. \tag{1.12}$$

Let $\mathbf{H}(\cdot)$ *be the entropy function using logarithm to base 2. Then,*

$$\sum_{i=1}^{k} \mathbf{H}(K[i]) \geq k - d.$$

PROOF. This is an application of Jensen's inequality. Let $q_x = \mathbb{P}(K = x)$. Then, $\mathbf{H}(K) = -\mathbb{E}\big(\log_2(q_K)\big)$ and we have

$$\begin{aligned} q_x \leq 2^{-(k-d)} \text{ for all } x &\implies \mathbb{E}[q_K] \leq 2^{-(k-d)} \implies \log_2\big(\mathbb{E}[q_K]\big) \leq -(k-d) \\ &\implies -\mathbb{E}[\log_2(q_K)] \geq k - d, \text{ by Jensen's inequality} \\ &\implies \mathbf{H}(K) \geq k - d. \end{aligned}$$

For discrete random variables Z, Z' on a common probability space, define the conditional entropy $\mathbf{H}(Z|Z') = \sum_z \mathbf{H}(Z|Z' = z)\mathbb{P}(Z' = z)$, where $\mathbf{H}(Z|Z' = z)$ denotes the entropy of the law of Z conditioned on $\{Z' = z\}$. Applying the chain rule for entropy on $K = (K[1], \ldots, K[k])$ gives us

$$\sum_{i=1}^{k} \mathbf{H}\big(K[i] \mid K[i-1], K[i-1], \ldots, K[1]\big) = \mathbf{H}(K) \geq k - d.$$

For any two discrete random variables Z, Z', the inequality $\mathbf{H}(Z|Z') \leq \mathbf{H}(Z)$ always holds. For a proof of this and the chain rule for conditional entropy, refer to Sections 2.2 and 2.5 from [**11**]. So the above inequality gives

$$\sum_{i=1}^{k} \mathbf{H}(K[i]) \geq \sum_{i=1}^{k} \mathbf{H}\big(K[i] \mid K[i-1], K[i-1], \ldots, K[1]\big) = \mathbf{H}(K) \geq k - d.$$

□

The entropy of a Bernoulli random variable can be used to bound the probability associated with it. This is the content of the lemma and corollary given below. We found the lemma on a math.stack.exchange webpage [**26**]. Relating the probability and entropy is useful since the conclusion of Theorem 1.5.1 is about the probabilities of a sub-string of the large string K. We already have a lower bound on the total entropy of all the bits in K.

LEMMA 1.5.3. *Let* $Z \sim Bernoulli(\theta)$*. Then,*

$$\mathbf{H}(Z) = -\theta \log_2(\theta) - (1-\theta)\log_2(1-\theta) \leq 2\sqrt{\theta(1-\theta)}, \text{ for } \theta \in [0, 1].$$

PROOF. Before we talk about $\mathbf{H}(Z)$, it will be useful to prove the following technical lemma: The equation

$$\frac{1-2x}{\sqrt{(1-x)x}} - \log_2(1-x) + \log_2(x) = 0$$

has exactly one solution on the interval $x \in (\frac{1}{2}, 1)$. To see this, begin by combining the logarithms, and moving them to the right hand side to get

$$\frac{1-2x}{\sqrt{(1-x)x}} = \log_2\left(\frac{1-x}{x}\right), \text{ which implies}$$
$$\frac{1-2x}{1-x} \cdot \frac{1-2x}{x} = \left[\log_2\left(\frac{1-x}{x}\right)\right]^2.$$

Note that we can square both sides without consequence, as both sides are negative on $(\frac{1}{2}, 1)$. Now substitute $y = \frac{1-x}{x}$. Note that $y - 1 = \frac{1-2x}{x}$, and $1 - \frac{1}{y} = \frac{1-2x}{1-x}$. Since $y(x) = \frac{1-x}{x}$ is a continuous bijection on $x \in (\frac{1}{2}, 1)$, we can equivalently show that the equation

$$\left(1 - \frac{1}{y}\right)(y-1) = [\log_2(y)]^2$$

has exactly one solution on the interval $y \in (0, 1)$. Multiplying by y on both sides gives us

$$(y-1)^2 = y[\log_2(y)]^2, \text{ which implies}$$
$$(y-1) = \sqrt{y}\log_2(y).$$

Note that we can safely take the positive square root on both sides, as $(y-1)$ and $\sqrt{y}\log_2(y)$ are both negative on $(0, 1)$. Let $g(y) = \sqrt{y}\log_2(y)$ and let $h(y) = (y-1)$.

The second derivative, $g''(y) = -\frac{1}{4}\log_2(y)y^{-3/2}$, is positive on (0,1], so $g(y)$ is convex on $(0, 1]$. It follow that $g(y)$ intersects the linear $h(y)$ at most twice on $(0, 1]$. One of those intersections is at $y = 1$, and since $g(\frac{1}{64}) > h(\frac{1}{64})$ and $g(\frac{1}{4}) < h(\frac{1}{4})$, the other intersection does in fact exist. So g and h intersect exactly once on $(0, 1)$. This proves the technical lemma.

Now we prove Lemma 1.5.3. We define the function $f(\theta)$ on $[0, 1]$ by

$$f(\theta) = 2\sqrt{\theta(1-\theta)} + \theta\log_2(\theta) + (1-\theta)\log_2(1-\theta)$$

where $f(0) = f(1) := 0$. Our goal is to show that $f(\theta) \geq 0$ for all $\theta \in [0,1]$. Note that as f is symmetric about $\theta = \frac{1}{2}$, it is enough the show that $f(\theta) \geq 0$ for all $\theta \in [\frac{1}{2}, 1]$. In order to do this, we will analyze the derivative of $f(\theta)$:

$$f'(\theta) = \frac{1-2\theta}{\sqrt{(1-\theta)\theta}} - \log_2(1-\theta) + \log_2(\theta).$$

By our technical lemma, the derivative has exactly one root in $(\frac{1}{2}, 1)$. This, along with the fact that $f(\frac{1}{2}) = f(1) = 0$ means that $f(\theta) \geq 0$ on $[\frac{1}{2}, 1]$ or $f(\theta) \leq 0$ on $[\frac{1}{2}, 1]$. By checking that $f(\frac{3}{4}) > 0$, we determine that it is the former. □

COROLLARY 1.5.2. *If* $\mathbf{H}(Z) = \epsilon$ *where* $Z \sim$ *Bernoulli*(θ) *just as above, then*

$$|\theta - 1/2| \leq \frac{\sqrt{1-\epsilon^2}}{2}.$$

PROOF. We get this by applying the above lemma and then completing the squares.

$$\epsilon \leq 2\sqrt{\theta(1-\theta)} \implies \theta^2 - \theta + \epsilon^2/4 \leq 0 \implies (\theta - 1/2)^2 \leq (1-\epsilon^2)/4 \implies |\theta - 1/2| \leq \sqrt{(1-\epsilon^2)/4}.$$

□

We now essentially prove Theorem 1.5.1 under the assumption that there is only one probe. We say essentially since the bound obtained here is slightly better than what Theorem 1.5.1 gives us for a single probe.

LEMMA 1.5.4. *Assume that a random string* $K \in \{0,1\}^k$ *satisfies*

$$\mathbb{P}(K = x) \leq \frac{1}{2^{k-d}} \text{ for all } x \text{ and some } d \geq 0. \tag{1.13}$$

Let P *be uniformly and independently chosen from* $\{1, \ldots, k\}$. *Then for* $Y = K[P]$ *and* $d \leq k$, *we have*

$$\mathbb{E}\left[\mathbb{P}(Y = 0 \mid P)^2 + \mathbb{P}(Y = 1 \mid P)^2\right] \leq 1 - \frac{1}{2}\left(1 - \frac{d}{k}\right)^2.$$

PROOF. Let $f(\theta) = \theta^2 + (1-\theta)^2$. Then, observe that whenever $|\theta - 1/2| \leq \delta$,

$$f(\theta) \leq 2\delta^2 + 1/2. \tag{1.14}$$

This is because $f(\theta) = \theta^2 + (1-\theta)^2 = 2(\theta - 1/2)^2 + 1/2$ is an increasing function of $|\theta - 1/2|$. Define $\mathrm{Ent}_i = \mathbf{H}(K[i])$. Then, applying lemma 1.5.2 we get

$$\mathbb{E}(\mathrm{Ent}_P) = \frac{1}{k}\sum_{i=1}^{k} \mathbf{H}(K[i]) \geq \frac{1}{k}(k-d) = 1 - \frac{d}{k}. \tag{1.15}$$

Applying corollary 1.5.2 to the random variable $K[P]$ conditioned on P, we get

$$\Big|\mathbb{P}\left(\, K[P] = 1 \mid P \,\right) - 1/2\Big| \leq \frac{1}{2}\sqrt{1 - \mathrm{Ent}_P^2}. \tag{1.16}$$

So,

$$\begin{aligned}
\mathbb{E}\Big[\mathbb{P}\left(\, Y = 0 \mid P \,\right)^2 + \mathbb{P}\left(\, Y = 1 \mid P \,\right)^2\Big] &= \mathbb{E}\Big[f\big(\mathbb{P}\left(\, K[P] = 1 \mid P \,\right)\big)\Big] \\
&\leq \mathbb{E}\left[2\left(\frac{1}{2}\sqrt{1 - \mathrm{Ent}_P^2}\right)^2 + \frac{1}{2}\right], \text{by (1.14) and (1.16)} \\
&= \mathbb{E}\left[1 - \frac{1}{2}\mathrm{Ent}_P^2\right] \\
&\leq 1 - \frac{1}{2}\left(\mathbb{E}\big[\mathrm{Ent}_P\big]\right)^2, \text{ by Jensen's inequality} \\
&\leq 1 - \frac{1}{2}\left(1 - \frac{d}{k}\right)^2, \text{ by (1.15).}
\end{aligned}$$

□

We will eventually prove Theorem 1.5.1 by applying Lemmas 1.5.5 and 1.5.6 below. Lemma 1.5.5 is an application of Lemma 1.5.4 given above.

Notation: let $Y^t = (K[P_1], K[P_2], \ldots, K[P_t])$ where $P_i's$ are the probes as usual. Let K^t denote the random string obtained by deleting the bits $K[P_1], K[P_2], \ldots, K[P_t]$ from the string K. Let $Y_i = K[P_i]$ and $\mathscr{P}^t = (P_1, P_2, \ldots, P_t)$. Also the define the following deterministic functions.

$$Q_t\left(b, y, \mathscr{p}, p\right) = \mathbb{P}\left(K[P_{t+1}] = b \mid Y^t = y, \mathscr{P}^t = \mathscr{p}, P_{t+1} = p\right) \text{ and} \tag{1.17}$$

$$P_t\left(y, \mathscr{p}\right) = \mathbb{P}\left(Y^t = y \mid \mathscr{P}^t = \mathscr{p}\right). \tag{1.18}$$

Note that $K, P_i, Y^t, \mathscr{P}^t$ are random and $Q_t(\cdot,\cdot,\cdot,\cdot), P_t(\cdot,\cdot)$ are not. However, when we apply the functions Q_t and P_t on random quantities, we get random variables. So, $Q_t(Y_{t+1}, Y^t, \mathscr{P}^t, P_{t+1})$ and $P_t(Y^t, \mathscr{P}^t)$ are random.

LEMMA 1.5.5. *Fix $n \in \mathbb{N}$ and let $t \leq n$. Assume that $\mathbb{P}(K = x) \leq 2^{-(k-d)}$ for all x and some $d \geq 0$. Let $\gamma \geq 2^{-(k-d-n)}$ .Then on the event $P_t(Y^t, \mathscr{P}^t) > \gamma$ the following is true.*

$$\mathbb{E}\Big[Q_t(Y_{t+1}, Y^t, \mathscr{P}^t, P_{t+1}) \mid P_t(Y^t, \mathscr{P}^t)\Big] \leq 1 - \frac{1}{2}\left(1 - \frac{d - \log_2(\gamma)}{k-n}\right)^2.$$

PROOF. We will prove this by applying Lemma 1.5.4 to the string K^t. Fix $Y^t = y$ and $\mathscr{P}^t = \mathscr{p}$ such that $\mathbb{P}\big(Y^t = y \mid \mathscr{P}^t = \mathscr{p}\big) = P_t(y, \mathscr{p}) \geq \gamma$. Let $\mathscr{p} = (p_1, \ldots, p_t)$ and $y = (y_1, \ldots, y_t)$, then

$$\begin{aligned}\mathbb{P}\left\{K^t = x \mid Y^t = y, \mathscr{P}^t = \mathscr{p}\right\} &= \frac{\mathbb{P}\big(K^t = x, Y^t = y, \mathscr{P}^t = \mathscr{p}\big)}{\mathbb{P}\big(Y^t = y, \mathscr{P}^t = \mathscr{p}\big)} \\ &= \frac{\mathbb{P}\big(K^t = x, K[p_1] = y_1, \ldots, K[p_t] = y_t\big)\mathbb{P}\big(\mathscr{P}^t = \mathscr{p}\big)}{P_t\big(y, \mathscr{p}\big)\mathbb{P}\big(\mathscr{P}^t = \mathscr{p}\big)}.\end{aligned}$$

$K^t = x, K[p_1] = y_1, \ldots, K[p_t] = y_t$ uniquely determines K. Say $K = z$. Then, continuing from above we get

$$\begin{aligned}\mathbb{P}\left\{K^t = x \mid Y^t = y, \mathscr{P}^t = \mathscr{p}\right\} &= \frac{\mathbb{P}(K = z)}{P_t(y, \mathscr{p})} \leq \frac{\mathbb{P}(K = z)}{\gamma} \\ &\leq \frac{2^{-(k-d)}}{\gamma} = \frac{1}{2^{(k-t)-(d-\log_2(\gamma)-t)}}.\end{aligned}$$

If $\gamma \geq 2^{-(k-d-n)}$, then $d - \log_2(\gamma) - t \leq k - t$. So we can now apply Lemma 1.5.4 to K^t and the probe P_{t+1} to get

$$\begin{aligned}&\mathbb{E}\Big[Q_t\big(0, Y^t, \mathscr{P}^t, P_{t+1}\big)^2 + Q_t\big(1, Y^t, \mathscr{P}^t, P_{t+1}\big)^2 \Big| \mathscr{P}^t = \mathscr{p}, Y^t = y\Big] \\ =&\mathbb{E}\Big[\mathbb{P}\big(K[P_{t+1}] = 0 \mid P_{t+1}, \mathscr{P}^t, Y^t\big)^2 + \mathbb{P}\big(K[P_{t+1}] = 1 \mid P_{t+1}, \mathscr{P}^t, Y^t\big)^2 \Big| \mathscr{P}^t = \mathscr{p}, Y^t = y\Big] \\ \leq& 1 - \frac{1}{2}\left(1 - \frac{d - \log_2(\gamma) - t}{k-t}\right)^2 \\ \leq& 1 - \frac{1}{2}\left(1 - \frac{d - \log_2(\gamma)}{k-n}\right)^2, \text{ since } t \leq n \text{ and } t \geq 0. \qquad (1.19)\end{aligned}$$

Now note that

$$\begin{aligned}
&\mathbb{E}\Big[Q_t(Y_{t+1}, Y^t, \mathscr{P}^t, P_{t+1}) \mid Y^t = y, \mathscr{P}^t = \mathscr{p}\Big] \\
&= \sum_p \sum_{b=0}^{1} Q_t(b, y, \mathscr{p}, p)\mathbb{P}\big(Y_{t+1} = b, P_{t+1} = p \mid Y^t = y, \mathscr{P}^t = \mathscr{p}\big) \\
&= \sum_p \sum_{b=0}^{1} Q_t(b, y, \mathscr{p}, p)\mathbb{P}\big(Y_{t+1} = b \mid Y^t = y, \mathscr{P}^t = \mathscr{p}, P_{t+1} = p\big)\mathbb{P}\big(P_{t+1} = p \mid Y^t = y, \mathscr{P}^t = \mathscr{p}\big) \\
&= \sum_p \sum_{b=0}^{1} Q_t(b, y, \mathscr{p}, p)^2\mathbb{P}\big(P_{t+1} = p \mid Y^t = y, \mathscr{P}^t = \mathscr{p}\big)
\end{aligned}$$

(1.20)

$$= \mathbb{E}\Big[Q_t\big(0, Y^t, \mathscr{P}^t, P_{t+1}\big)^2 + Q_t\big(1, Y^t, \mathscr{P}^t, P_{t+1}\big)^2 \Big| \mathscr{P}^t = \mathscr{p}, Y^t = y\Big].$$

From (1.19) and (1.20) above we can conclude that on the event $P_t(Y^t, \mathscr{P}^t) \geq \gamma$,

$$\mathbb{E}\Big[Q_t(Y_{t+1}, Y^t, \mathscr{P}^t, P_{t+1}) \mid Y^t, \mathscr{P}^t\Big] \leq 1 - \frac{1}{2}\left(1 - \frac{d - \log_2(\gamma)}{k - n}\right)^2.$$

So, we finally have that on the event $P_t(Y^t, \mathscr{P}^t) \geq \gamma$,

$$\begin{aligned}
\mathbb{E}\Big[Q_t(Y_{t+1}, Y^t, \mathscr{P}^t, P_{t+1}) \mid P_t(Y^t, \mathscr{P}^t)\Big] &= \mathbb{E}\left[\mathbb{E}\Big[Q_t(Y_{t+1}, Y^t, \mathscr{P}^t, P_{t+1}) \mid Y^t, \mathscr{P}^t\Big] \;\Big|\; P_t(Y^t, \mathscr{P}^t)\right] \\
&\leq \mathbb{E}\left[1 - \frac{1}{2}\left(1 - \frac{d - \log_2(\gamma)}{k - n}\right)^2 \;\Big|\; P_t(Y^t, \mathscr{P}^t)\right] \\
&= 1 - \frac{1}{2}\left(1 - \frac{d - \log_2(\gamma)}{k - n}\right)^2.
\end{aligned}$$

□

The following lemma is from [**24**]. The lemma gives a bound on $\mathbb{E}[Z_n]$ for random variables $Z_1, Z_2, \ldots$ under the assumption that for fixed Z_t, on average the value of Z_{t+1} is smaller than Z_t by a factor which depends on Z_t.

LEMMA 1.5.6. *Let $f : [0, \infty) \to [0, 1]$ be an increasing function. Suppose that $\{Z_t\}_{t \geq 0}$ are non-negative random variables with $Z_0 = L_0$. Denote $L_n = \mathbb{E}[Z_n]$. Assume that $\mathbb{E}[Z_{t+1} \mid Z_t] \leq (1 - f(Z_t))Z_t$ for all t. Then for every $t \geq \int_\delta^{L_0} \frac{2}{zf(z/2)}dz$, we have $L_t \leq \delta$.*

PROOF. This is part (iii) of Lemma 11 in [**24**]. □(1.5.1).

PROOF. (Theorem 1.5.1) Let $P_t(\cdot,\cdot)$ and $Q_t(\cdot,\cdot,\cdot,\cdot)$ be as defined in (1.17) and (1.18). Then, note that

$$\begin{aligned}
\mathbb{E}[P_t(Y^t,\mathscr{P}^t)] &= \sum_{y,p} P_t(y,p)\mathbb{P}\big(Y^t=y,\mathscr{P}^t=p\big)\\
&= \sum_{y,p} P_t(y,p)\mathbb{P}\big(Y^t=y\mid\mathscr{P}^t=p\big)\mathbb{P}\big(\mathscr{P}^t=p\big)\\
&= \sum_{p}\sum_{y} P_t(y,p)^2\big(\mathscr{P}^t=p\big)\\
&= \mathbb{E}\Big[\sum_{y\in\{0,1\}^t}\mathbb{P}\big(Y^t=y\mid\mathscr{P}^t\big)^2\Big]. \qquad (1.21)
\end{aligned}$$

This is the quantity we want to bound. Let n be as in the hypothesis. For $2^{-(k-d-n)}\le\gamma\le 1$ set

$$f(x)=\begin{cases}\frac{1}{2}\left(1-\frac{d-\log_2(\gamma)}{k-n}\right)^2 & \text{if } x\ge\gamma\\ 0 & \text{if } x<0\end{cases}$$

We will choose γ later. For $1\le t\le n$, set

$$Z_t = P_t(Y^t,\mathscr{P}^t).$$

Define $Z_0=1$ and for $t\ge n+1$ define

$$Z_t=(1-f(Z_{t-1}))Z_{t-1}.$$

Note that from the above, $\mathbb{E}[Z_t\mid Z_{t-1}]\le(1-f(Z_{t-1}))Z_{t-1}$ whenever $t\ge n+1$. For $2\le t\le n$, note that if $y=(y_1,\dots,y_{t-1})$, $p'=(p_1,\dots,p_{t-1})$, $y'=(y_1,\dots,y_{t-1},b)$ and $p=(p_1,\dots,p_{t-1},p)$ then

$$\begin{aligned}
P_t(y',p') &= \frac{\mathbb{P}(Y^t=y',\mathscr{P}^t=p')}{\mathbb{P}(\mathscr{P}^t=p')}\\
&= \frac{\mathbb{P}(Y_t=b,Y^{t-1}=y,\mathscr{P}^{t-1}=p,P_t=p)}{\mathbb{P}(Y^{t-1}=y,\mathscr{P}^{t-1}=p,P_t=p)}\cdot\frac{\mathbb{P}(Y^{t-1}=y,\mathscr{P}^{t-1}=p,P_t=p)}{\mathbb{P}(\mathscr{P}^{t-1}=p,P_t=p)}\\
&= Q_{t-1}(b,y,p,p)\cdot\frac{\mathbb{P}(Y^{t-1}=y,\mathscr{P}^{t-1}=p,P_t=p)}{\mathbb{P}(\mathscr{P}^{t-1}=p,P_t=p)}. \qquad (1.22)
\end{aligned}$$

Since $P_1, \ldots, P_t$ are independent of K, we get

$$\begin{aligned}\frac{\mathbb{P}(Y^{t-1} = y, \mathscr{P}^{t-1} = \mathscr{p}, P_t = p)}{\mathbb{P}(\mathscr{P}^{t-1} = \mathscr{p}, P_t = p)} &= \frac{\mathbb{P}(K[p_1] = y_1, \ldots, K[p_{t-1}] = y_{t-1})\mathbb{P}(\mathscr{P}^{t-1} = \mathscr{p}, P_t = p)}{\mathbb{P}(\mathscr{P}^{t-1} = \mathscr{p}, P_t = p)} \\ &= \frac{\mathbb{P}(K[p_1] = y_1, \ldots, K[p_{t-1}] = y_{t-1})\mathbb{P}(\mathscr{P}^{t-1} = \mathscr{p})}{\mathbb{P}(\mathscr{P}^{t-1} = \mathscr{p})} \\ &= \frac{\mathbb{P}(Y^{t-1} = y, \mathscr{P}^{t-1} = \mathscr{p})}{\mathbb{P}(\mathscr{P}^{t-1} = \mathscr{p})} = P_{t-1}(y, \mathscr{p}).\end{aligned}$$

So, from (1.22) we get

$$P_t(y', \mathscr{p}') = Q_{t-1}(b, y, \mathscr{p}, p)P_{t-1}(y, \mathscr{p}).$$

From above, we get that for $2 \leq t \leq n$,

$$Z_t = Q_{t-1}(Y_t, Y^{t-1}, \mathscr{P}^{t-1}, P_t) \cdot Z_{t-1}. \tag{1.23}$$

From Lemma 1.5.5, we see that whenever $P_{t-1}(Y^{t-1}, \mathscr{P}^{t-1}) = Z_{t-1} \geq \gamma$ and $2 \leq t \leq n$,

$$\mathbb{E}\big[Q_{t-1}(Y_t, Y^{t-1}, \mathscr{P}^{t-1}, P_t) \mid Z_{t-1}\big] \leq 1 - \frac{1}{2}\left(1 - \frac{d - \log_2(\gamma)}{k - n}\right)^2 = (1 - f(Z_{t-1})).$$

If $Z_{t-1} < \gamma$ then, since $Q_{t-1}(Y_t, Y^{t-1}, \mathscr{P}^{t-1}, P_t) \leq 1$, we have

$$\mathbb{E}\big[Q_{t-1}(Y_t, Y^{t-1}, \mathscr{P}^{t-1}, P_t) \mid Z_{t-1}\big] \leq 1 - 0 = 1 - f(Z_{t-1}).$$

So, we have shown that for $2 \leq t \leq n$,

$$\mathbb{E}\big[Z_t \mid Z_{t-1}\big] = \mathbb{E}\big[Q_{t-1}(Y_t, Y^{t-1}, \mathscr{P}^{t-1}, P_t) \mid Z_{t-1}\big] \cdot Z_{t-1} \leq (1 - f(Z_{t-1})) \cdot Z_{t-1}.$$

For $t \geq n + 1$ the above is true by definition, as remarked previously. For $t = 1$, by Lemma 1.5.4,

$$\mathbb{E}[Z_1 \mid Z_0] = \mathbb{E}[Z_1] \leq 1 - \frac{1}{2}\left(1 - \frac{d}{k}\right)^2 \leq 1 - \frac{1}{2}\left(1 - \frac{d - \log_2(\gamma)}{k - n}\right)^2,$$

since $-\log_2(\gamma) \geq 0$ and $n \geq 0$. So, by applying Lemma 1.5.6, we get that for $t \geq \int_\delta^1 \frac{2}{zf(z/2)}dz$,

$$L_t = \mathbb{E}[Z_t] \leq \delta.$$

Choose $\delta = 2\gamma$ and choose γ to be a solution to the equation

$$n = \int_{2\gamma}^{1} \frac{2}{zf(z/2)} dz,$$

such that $2^{-(k-d-n)} \leq \gamma \leq 1$. If we simplify the above equation, we get

$$-\log_2(\gamma) = 1 + \frac{n}{4\ln 2}\left(1 - \frac{d - \log_2(\gamma)}{k-n}\right)^2. \tag{1.24}$$

When $\gamma = 1$,

$$\log_2(\gamma) + 1 + \frac{n}{4\ln 2}\left(1 - \frac{d - \log_2(\gamma)}{k-n}\right)^2 = 1 + \frac{n}{4\ln 2}\left(1 - \frac{d}{k-n}\right)^2 > 0.$$

Recall that from the hypothesis, $d + 1 + n/(4\ln 2) \leq k - n$. So, for $\gamma = 2^{-(k-d-n)}$,

$$\log_2(\gamma) + 1 + \frac{n}{4\ln 2}\left(1 - \frac{d - \log_2(\gamma)}{k-n}\right)^2 = -(k-d-n) + 1 \leq -n/4\ln 2 < 0.$$

This shows that γ indeed does exist. Next, note that for $2^{-(k-d-n)} \leq \gamma \leq 1$,

$$\left(1 - \frac{d - \log_2(\gamma)}{k-n}\right)^2 \leq 1.$$

So, from equation (1.24), we get $-\log(\gamma) \leq 1 + n/(4\ln 2)$. This implies that

$$\left(1 - \frac{d - \log_2(\gamma)}{k-n}\right)^2 \geq \left(1 - \frac{d + 1 + n/(4\ln 2)}{k-n}\right)^2.$$

If we use equation (1.24) once again, we get

$$-\log_2(\gamma) \geq 1 + \frac{n}{4\ln 2}\left(1 - \frac{d + 1 + n/(4\ln 2)}{k-n}\right)^2.$$

Recall that $\beta(k,d,n) = \frac{1}{8\ln 2}\left(1 - \frac{d+1+n/(4\ln 2)}{k-n}\right)^2$. Hence, from the above equation we get

$$\gamma \leq 2^{-1-2n\cdot\beta(k,d,n)}.$$

As remarked above, for $t \geq \int_{2\gamma}^{1} \frac{1}{zf(z/2)} dz = n$, we have $L_t \leq 2\gamma$. So in particular this holds for $t = n$. So we can now use (1.21) to conclude

$$\mathbb{E}\Big[\sum_{y \in \{0,1\}^n} \mathbb{P}\big(Y^n = y \mid \mathscr{P}^n\big)^2 \Big] = \mathbb{E}[Z_n] = L_n \leq 2\gamma \leq 2 \cdot 2^{-1-2n\cdot\beta(k,d,n)} = 2^{-2n\cdot\beta(k,d,n)}.$$

□

1.6. Randomness of the Bits Generated by Probing the Key Under Leakage

Note that the results of Section 1.5 are purely mathematical and are interesting on their own without any reference to cryptography. In this section we will be going back to our original setup of keys and leakage. That is, from now on we will use K to denote the key and we will use $L = \Phi(K)$ to denote the leakage. Until the final section, we will only look at the special case where the leakage function doesn't depend on the random oracle. What this means mathematically is that the probes and subprobes are truly random even when we condition on the leakage. In other words, this means that the probes and sub probe indices are independent of the leakage. We will solve the case where the leakage depends on random oracle calls by applying the preceding case. We will do this in the final section.

COROLLARY 1.6.1. *Let K be the key and $L = \Phi(K)$ the leakage. Let $c, P_1, \ldots, P_n$ and $\mathcal{S}$ be as in section 1.3; let $\mathscr{P}^n = (P_1, \ldots, P_n)$ and $c' \sim$ Bernoulli$(1/2)$. Assume that the random oracle call required to compute the probes and sub probes corresponding to c is not made by the adversary while deciding on $\Phi(\cdot)$. Assume that $d \geq 0$ with $d + 1 + n/(4\ln 2) \leq k - n$. Then on the event $1/|\Phi^{-1}(L)| \leq 1/2^{k-d}$, we have*

$$\mathbb{E}\Big[\big\| \mathscr{L}\big(c \mid L, \mathscr{P}^n, \mathcal{S}\big) - \mathscr{L}\big(c'\big) \big\|_{TV} \;\Big|\; L \Big] \leq 2^{-1-n\cdot\beta(k,d,n)}, \text{ where}$$

$$\beta(k,d,n) = \frac{1}{8\ln 2} \left(1 - \frac{d + 1 + n/(4\ln 2)}{k - n}\right)^2.$$

PROOF. Let $L = x$ and assume that $1/|\Phi^{-1}(x)| \leq 1/2^{k-d}$. Then, $\mathbb{P}(K = k \mid L = x) \leq 1/|\Phi^{-1}(L)| \leq 1/2^{k-d}$. From Corollary (1.5.1) it follows that,

$$\mathbb{E}\left[\left|\left|\mathscr{L}\big(c \mid L, \mathscr{P}^n, \mathcal{S}\big) - \mathscr{L}\big(c'\big)\right|\right|_{TV} \;\middle|\; L = x\right] \leq \mathbb{E}\left[\frac{1}{2}\sqrt{\sum_{y\in\{0,1\}^n} \mathbb{P}\left(Y^n = y \mid \mathscr{P}^n, L\right)^2} \;\middle|\; L = x\right]$$

$$\leq \frac{1}{2}\sqrt{\mathbb{E}\left[\sum_{y\in\{0,1\}^n} \mathbb{P}\left(Y^n = y \mid \mathscr{P}^n, L\right)^2 \;\middle|\; L = x\right]}, \text{ by Jensen's inequality}$$

$$\leq \frac{1}{2}\sqrt{2^{-2n\cdot\beta(k,d,n)}} = 2^{-1-n\cdot\beta(k,d,n)}.$$

where last inequality is obtained by applying theorem (1.5.1) to K conditioned on the event $L = x$. □

LEMMA 1.6.1. *Let Φ be a fixed leakage function. Then,*

$$\mathbb{E}\Big[1/\big|\Phi^{-1}(L)\big|\Big] \leq 1/2^{k-l}.$$

PROOF. For $i = 1, 2, \ldots, 2^l$ let $S_i = \Phi^{-1}(i)$, where the i on the right is interpreted to be in binary notation. Note that $\Phi^{-1}(L) = S_i$ if and only if $K \in S_i$. So, $\mathbb{P}\left(\Phi^{-1}(L) = S_i\right) = |S_i|/2^k$ and

$$\begin{aligned}\mathbb{E}\Big[1/\big|\Phi^{-1}(L)\big|\Big] &= \sum_{i=1}^{2^l} \mathbb{P}\left(\Phi^{-1}(L) = S_i\right)\frac{1}{|S_i|} = \sum_{i=1}^{2^l}\frac{|S_i|}{2^k}\frac{1}{|S_i|}\\ &= \frac{1}{2^{k-l}}.\end{aligned}$$

□

LEMMA 1.6.2. *Let $c, c', L, K, \mathscr{P}^n$ and $\mathcal{S}$ be as in corollary 1.6.1. Assume that the random oracle call required to compute the probes and sub probes corresponding to c is not made by the adversary while deciding on $\Phi(\cdot)$. Then for $l + 1 + n/(2\ln 2) \leq k - n$,*

$$\mathbb{E}\Big[\big|\big|\mathscr{L}(c|L, \mathscr{P}^n, \mathcal{S}) - \mathscr{L}(c')\big|\big|_{TV}\Big] \leq 2\cdot 2^{-n\cdot\alpha(k,l,n)}, \text{ where}$$

$$\alpha(k, l, n) = \frac{1}{8\ln 2}\left(1 - \frac{l + 1 + n/(2\ln 2)}{k - n}\right)^2.$$

PROOF. Let $\mathcal{I}_1 = \mathbb{1}_{\{1/|\Phi^{-1}(L)|\le 1/2^{k-d}\}}$ and $\mathcal{I}_2 = \mathbb{1}_{\{1/|\Phi^{-1}(L)|>1/2^{k-d}\}}$. Think of $\mathscr{L}(c|L,\mathscr{P}^n,\mathcal{S})$ as a vector function of $(L,\mathscr{P}^n,\mathcal{S})$. Then by triangle equality for total variation distance we have,

$$\begin{aligned}
&\mathbb{E}\Big[\big\|\mathscr{L}(c|L,\mathscr{P}^n,\mathcal{S})-\mathscr{L}(c')\big\|_{TV}\ \Big|L\Big]\\
=&\mathbb{E}\Big[\big\|(\mathcal{I}_1+\mathcal{I}_2)\mathscr{L}(c|L,\mathscr{P}^n,\mathcal{S})-(\mathcal{I}_1+\mathcal{I}_2)\mathscr{L}(c')\big\|_{TV}\ \Big|L\ \Big]\\
\le&\mathbb{E}\Big[\big\|\mathcal{I}_1\mathscr{L}(c|L,\mathscr{P}^n,\mathcal{S})-\mathcal{I}_1\mathscr{L}(c')\big\|_{TV}+\big\|\mathcal{I}_2\mathscr{L}(c|\Phi^{-1}(L),\mathscr{P}^n)-\mathcal{I}_2\mathscr{L}(c')\big\|_{TV}\ \Big|\ L\Big]. \qquad (1.25)
\end{aligned}$$

The first term above can be bounded by noting that when $\mathcal{I}_1 = 0$, the term is zero and when $\mathcal{I}_1 = 1$, we can use Corollary [1.6.1]. So we get

$$\mathbb{E}\Big[\big\|\mathcal{I}_1\mathscr{L}(c|L,\mathscr{P}^n,\mathcal{S})-\mathcal{I}_1\mathscr{L}(c')\big\|_{TV}\ \Big|\ L\Big]\le 2^{-1-n\cdot\beta(k,d,n)}, \qquad (1.26)$$

where $\beta(k,d,n)=\frac{1}{8\ln 2}\left(1-\frac{d+1+n/(4\ln 2)}{k-n}\right)^2$. Next, to bound the second term in (1.25), note that the total variation distance is always bounded by one and hence,

$$\begin{aligned}
\mathbb{E}\Big[\big\|\mathcal{I}_2\mathscr{L}(c|L,\mathscr{P}^n,\mathcal{S})-\mathcal{I}_2\mathscr{L}(c')\big\|_{TV}\ \Big|\ L\Big]&=\mathbb{E}\Big[\mathcal{I}_2\big\|\mathscr{L}(c|L,\mathscr{P}^n,\mathcal{S})-\mathscr{L}(c')\big\|_{TV}\ \Big|\ L\Big]\\
&\le\mathbb{E}\big[\mathcal{I}_2\ \big|\ L\big]. \qquad (1.27)
\end{aligned}$$

If we now apply $\mathbb{E}$ to (1.25) and use (1.26) and (1.27), we get

$$\begin{aligned}
\mathbb{E}\Big[\big\|\mathscr{L}(c|L,\mathscr{P}^n,\mathcal{S})-\mathscr{L}(c')\big\|_{TV}\Big]&\le 2^{-1-n\cdot\beta(k,d,n)}+\mathbb{E}[\mathcal{I}_2]\\
&=2^{-1-n\cdot\beta(k,d,n)}+\mathbb{P}\big(1/|\Phi^{-1}(L)|>1/2^{k-d}\big)\\
&\le 2^{-1-n\cdot\beta(k,d,n)}+2^{k-d}\mathbb{E}\big[1/|\Phi^{-1}(L)|\big],\ \text{by Markov's inequality}\\
&=2^{-1-n\cdot\beta(k,d,n)}+2^{-(d-l)}. \qquad (1.28)
\end{aligned}$$

This is true for any $d\ge 0$ with $d+1+n/(4\ln 2)\le k-n$. Choose $d=l+n/4\ln 2$. Then, since $\beta(k,d,n)\le 1/8\ln 2\le 1/4\ln 2$, we have $2^{-(d-l)}=2^{-n/4\ln 2}\le 2^{-n\beta(k,d,n)}$. So from (1.28) we get

$$\begin{aligned}
\mathbb{E}\Big[\big\|\mathscr{L}(c|L,\mathscr{P}^n,\mathcal{S})-\mathscr{L}(c')\big\|_{TV}\Big]&\le 2^{-1-n\cdot\beta(k,d,n)}+2^{-n\cdot\beta(k,d,n)}\\
&\le 2\cdot 2^{-n\cdot\beta(k,l+n/4\ln 2,n)}=2\cdot 2^{-n\cdot\alpha(k,l,n)}.
\end{aligned}$$

□

This shows that on average, the bit c is very close to a Bernoulli(1/2) random variable, even when the leakage, probes and the sub probe indices are known. We will now need an analogous result for a sequence of bits $(c_1, \ldots, c_t)$. This is achieved by the next theorem which shows that even conditioned on $(c_1, \ldots, c_{t-1})$, the next bit c_t is very close to a Bernoulli(1/2) bit. We prove this theorem by thinking of the bits $c_1, \ldots, c_{t-1}$ as an additional $t-1$ bits of leaked information and applying Lemma 1.6.2.

THEOREM 1.6.1. *Let $c_1, c_2, \ldots, c_{h+1}$ be bits generated in a similar fashion to c. Let $\mathscr{P}_i^n$ be the vector of probes associated with the bit c_i and let $\mathcal{S}_i$ be the corresponding sub probe indices. Let $\overline{\mathscr{P}} = (\mathscr{P}_1^n, \mathscr{P}_2^n, \ldots, \mathscr{P}_{h+1}^n)$ and $\overline{\mathcal{S}} = (\mathcal{S}_1, \ldots, \mathcal{S}_{h+1})$. Assume that the random oracle involved in generating $c_i's$ is applied on different inputs so that the $\mathscr{P}_i^n$ are independent of each other for various $i's$ and also independent of the key. Also assume that the adversary doesn't make any of the random oracle calls necessary to compute the $c_i's$ while choosing Φ. Let c' be a Bernoulli(1/2) random variables. Then for $l + h + n/(2\ln 2) + 1 \leq k - n$,*

$$\mathbb{E}\Big[\big\|\mathscr{L}\big(c_{h+1} \mid c_h, \ldots, c_1, L, \overline{\mathscr{P}}, \overline{\mathcal{S}}\big) - \mathscr{L}(c')\big\|_{TV}\Big] \leq 2 \cdot 2^{-n\cdot\alpha(k,l+h,n)}, \text{ where}$$

$$\alpha(k, a, n) = \frac{1}{8\ln 2}\left(1 - \frac{a + 1 + n/(2\ln 2)}{k - n}\right)^2.$$

PROOF. Let $\overline{\mathscr{P}}' = (\mathscr{P}_1^n, \mathscr{P}_2^n, \ldots, \mathscr{P}_h^n)$ and $\overline{\mathcal{S}}' = (\mathcal{S}_1, \ldots, \mathcal{S}_h)$. Condition on $\{\overline{\mathscr{P}}' = A, \overline{\mathcal{S}}' = B\}$ and define the function $\widetilde{\Phi}_{A,B} : \{0,1\}^k \to \{0,1\}^{l+h}$ as follows.

$$\widetilde{\Phi}_{A,B}(x) = (\Phi(x), c_1, c_2, \ldots, c_h)$$

where $(c_1, \ldots, c_h)$ are the the bits generated when $\overline{\mathscr{P}}' = A$, $\overline{\mathcal{S}} = B$ and x is chosen as the key. Conditioning on $\overline{\mathscr{P}}', \overline{\mathcal{S}}'$ doesn't change the distribution of $K, \mathscr{P}_{h+1}^n$ and $\mathcal{S}_{h+1}$. So we can apply Lemma 1.6.2 by replacing Φ with $\widetilde{\Phi}_{A,B}$ to get

$$\mathbb{E}\left[\left\|\mathscr{L}\left(c_{h+1} \mid \widetilde{\Phi}_{\overline{\mathscr{P}}',\overline{\mathcal{S}}'}(K), \overline{\mathscr{P}}, \overline{\mathcal{S}}\right) - \mathscr{L}(c'_{h+1})\right\|_{TV} \;\middle|\; \overline{\mathscr{P}}' = A, \overline{\mathcal{S}}' = B\right] \leq 2 \cdot 2^{-n\cdot\alpha(k,l+h,n)}.$$

The right hand side of the above inequality doesn't depend on the value of $\overline{\mathscr{P}}'$ and $\overline{\mathcal{S}}'$. So we can an take expectation over $\overline{\mathscr{P}}', \overline{\mathcal{S}}'$ to get

$$\mathbb{E}\left[\left\| \mathscr{L}\left(c_{h+1} \mid \widetilde{\Phi}_{\overline{\mathscr{P}}', \overline{\mathcal{S}}'}(K), \overline{\mathscr{P}}, \overline{\mathcal{S}} \right) - \mathscr{L}(c'_{h+1}) \right\|_{TV} \right] \leq 2 \cdot 2^{-n \cdot \alpha(k,l+h,n)}.$$

This completes the proof since $\widetilde{\Phi}_{\overline{\mathscr{P}}', \overline{\mathcal{S}}'}(K) = (L, c_1, \ldots, c_h)$. □

In the next section we will argue that since the bits we have generated in our algorithm are very close to i.i.d. Bernoulli(1/2) random bits, it must be that our algorithm/shuffle is close to the Thorp shuffle in total variation distance.

1.7. Comparison with the Thorp shuffle

Let $M_1, M_2, \ldots, M_q$ be messages (queries). Let $\mathrm{Al}_t(M_{i_1}, M_{i_2}, \ldots, M_{i_o})$ be the result of applying t steps of our algorithm to messages $M_{i_1}, M_{i_2}, \ldots, M_{i_o}$. Let $\mathrm{Th}_t(M_{i_1}, M_{i_2}, \ldots, M_{i_o})$ be the result of applying t steps of the Throp shuffle to messages $M_{i_1}, \ldots, M_{i_o}$. The Thorp shuffle is defined inductively as follow. For a message M, say $\mathrm{Th}_{t-1}(M) = b_1 b_2 \ldots b_m$. Let M' be the message such that $\mathrm{Th}_{t-1}(M') = (b_1 \oplus 1) b_2 \ldots b_m$. We now take a bit $c' \sim$ Bernoulli(1/2) and set $\mathrm{Th}_t(M) = b_2 b_3 \ldots b_m (b_1 \oplus c')$ and $\mathrm{Th}_t(M') = b_2 b_3 \ldots b_m (b_1 \oplus 1 \oplus c')$. The various Bernoulli bits we use are taken to be independent of each other. Our algorithm is actually a modified Thorp shuffle in which instead of using i.i.d. Bernoulli bits, we use bits that are generated by probing a key. The Thorp shuffle was first used for format preserving encryption in [**25**]. Our aim is to bound the total variation distance between the outcome of our algorithm and that of the Thorp shuffle. This is the content of Lemma 1.7.1. We will need the notion of optimal coupling in order to prove this theorem. So we digress a bit and layout some definitions and results below.

Definition. For probability distributions μ and ν, we say that a pair of random variables (X, Y) with some joint distribution is a coupling if $\mu(\cdot) = \mathbb{P}\big(X = \cdot \big)$ and $\nu(\cdot) = \mathbb{P}\big(Y = \cdot \big)$. If in addition we have $\|\mu - \nu\|_{TV} = \mathbb{P}\big(X \neq Y \big)$, then it is called an optimal coupling.

It can be shown that for discrete probability distributions, an optimal coupling always exists. It is

also true that for any two discrete probability distributions μ and ν,

$$\|\mu - \nu\|_{TV} = \inf_{(X,Y) \text{ is a coupling}} \mathbb{P}\big(X \neq Y\big). \tag{1.29}$$

A proof of these two statements can be found in Section 4.2 of [**19**].

LEMMA 1.7.1. *Let L be the leakage. Let $\overline{\mathscr{P}}$ be the tuple consisting of all the probes used to compute $\mathrm{Al}_T(M_1, \ldots, M_q)$ and let $\overline{\mathcal{S}}$ be the tuple consisting of all the associated sets of sub probe indices. Let $\mathscr{L}$ be the distribution of $\mathrm{Al}_T(M_1, \ldots, M_q)$ conditioned on $L, \overline{\mathscr{P}}, \overline{\mathcal{S}}$ and let $\mathscr{L}'$ be the distribution of $\mathrm{Th}_T(M_1, \ldots, M_q)$. Assume that the adversary hasn't made any of the random oracle calls that are needed to compute $\overline{\mathscr{P}}$ and $\overline{\mathcal{S}}$. Then for $l + qT + n/(2\ln 2) \leq k - n$,*

$$\mathbb{E}\|\mathscr{L} - \mathscr{L}'\| \leq 2qT \cdot 2^{-n \cdot \alpha(k, l+qT-1, n)},$$

where k is the length of the key, l is the length of the leakage, n is the number of probes used to compute a single bit in our algorithm and

$$\alpha(k, a, n) = \frac{1}{8\ln 2}\left(1 - \frac{a + 1 + n/(2\ln 2)}{k - n}\right)^2.$$

PROOF. Fix $L, \overline{\mathscr{P}}, \overline{\mathcal{S}}$. Let $C(l,t)$ be the random bit used at step t of our algorithm for the message M_l. That is, if $\mathrm{Al}_{t-1}(M_l) = b_1 b_2 \ldots b_m$ then $\mathrm{Al}_t(M_l) = b_2 \ldots b_m (b_1 \oplus C(l,t))$. Similarly, let $C'(l,t)$ be the random bit used at step t of the Thorp shuffle for the message M_l. Let $C_1, C_2, \ldots, C_{qT}$ be a reordering of $\{C(l,t) \ : \ 1 \leq l \leq q, 1 \leq t \leq T\}$ such that for $C_i = C(l_1, t_1)$ and $C_j = C(l_2, t_2)$, we have $i \leq j$ if and only if either $l_1 < l_2$, or $l_1 = l_2$ and $t_1 \leq t_2$. That is we order $\{C(l,t) \ : \ 1 \leq l \leq q, 1 \leq t \leq T\}$ according to the dictionary order on (l,t). Let $C'_1, C'_2, \ldots, C'_{qT}$ be a similar ordering for $\{C'(l,t) \ : \ 1 \leq l \leq q, 1 \leq t \leq T\}$. With this notation, note that $\mathrm{Al}_T(M_1, \ldots, M_q)$ is a function of $C_1, \ldots, C_{qT}$. Say,

$$\mathrm{Al}_T(M_1, \ldots, M_q) = f(C_1, \ldots, C_{qT}).$$

Then, note that since Th_T is obtained in the same way as Al_T, but using the C' bits, we have

$$\mathrm{Th}_T(M_1, \ldots, M_q) = f(C'_1, \ldots, C'_{qT}).$$

Let $\widetilde{\mathscr{L}}(\cdot)$ denote the distribution of a random variable conditioned on $L, \overline{\mathscr{P}}, \overline{\mathcal{S}}$. Then, we have

$$\begin{aligned}\|\mathscr{L} - \mathscr{L}'\|_{TV} &= \big\|\widetilde{\mathscr{L}}(f(C_1, \ldots, C_{qT})) - \mathscr{L}(f(C'_1, \ldots, C'_{qT}))\big\|_{TV} \\ &\leq \big\|\widetilde{\mathscr{L}}(C_1, \ldots, C_{qT}) - \mathscr{L}(C'_1, \ldots, C'_{qT})\big\|_{TV}.\end{aligned}$$

We get the above inequality from the following equivalent definition of total variation distance. For any two random variables X, Y supported on a finite set Ω, we have $\|\mathscr{L}(X) - \mathscr{L}(Y)\|_{TV} = \max_{A\subseteq\Omega}[\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)]$. Applying this, we get $\|\mathscr{L}(f(X)) - \mathscr{L}(f(Y))\|_{TV} = \max_{A\subseteq\Omega}[\mathbb{P}(X \in f^{-1}(A)) - \mathbb{P}(Y \in f^{-1}(A))] \leq \max_{A\in\Omega}\|\mathscr{L}(X) - \mathscr{L}(Y)\|_{TV}$. For a proof of equivalence of the above definition of total variation distance with the one we gave before this, refer Section 4.1 in [**19**].

For a given i, say there exists $i' < i$ such that $C_i = C(l, t)$ and $C_{i'} = C(l', t)$. Also say $\mathrm{Th}_{t-1}(M_{l'}) = b_1 b_2 \ldots b_m$ and $\mathrm{Th}_{t-1}(M_l) = (b_1 \oplus 1) b_2 \ldots b_m$. If such an i' exists, then we say C_i is an *old* bit. If no such i' exists we say C_i is a *fresh* bit. Loosely speaking, if the bit used at step t for message M_l is the same as the bit used at step t for the message $M_{l'}$ where $l' < l$ then we say the bit $C(l, t)$ is old. Define a bit C'_i to be old/fresh in the same way. We are now going to couple the C_is and the C'_is inductively as follows. Choose (C_1, C'_1) to be the optimal coupling of the distributions $\mathscr{L}(C_1 \mid \overline{\mathscr{P}}, \overline{\mathcal{S}}, L)$ and $\mathscr{L}(C'_1)$. Next, assuming $(C_1, \ldots, C_{i-1})$ and $(C'_1, \ldots, C'_{i-1})$ have been coupled, we will couple C_i and C'_i. First, we couple these two random variable arbitrarily on the event $(C_1, \ldots, C_{i-1}) \neq (C'_1, \ldots, C'_{i-1})$. Then, for any binary string $(a_1, \ldots, a_{i-1})$ we consider the event $(C_1, \ldots, C_{i-1}) = (C'_1, \ldots, C'_{i-1}) = (a_1, \ldots, a_{i-1})$. If C_i is an old bit, then on the preceding event, C'_i must also be an old bit. This is because if $C_i = C(l, t)$ and $C'_i = C'(l, t)$ then on the event $(C_1, \ldots, C_{i-1}) = (C'_1, \ldots, C'_{i-1})$, we have $\mathrm{Al}_{t-1}(M_{l'}) = \mathrm{Th}_{t-1}(M_{l'})$ for every $l < l'$. By a similar reasoning, if C'_i is an old bit then on the above event, C_i must also be an old bit. Old bits have already been coupled. So assume that both C_i and C'_i are fresh bits. Choose (C_i, C'_i) according to the optimal coupling of the distributions $\mathscr{L}\big(C_i \mid C_j = a_j \text{for } 1 \leq j \leq i-1, \overline{\mathscr{P}}, \overline{\mathcal{S}}, L\big)$ and $\mathscr{L}\big(C'_i \mid C'_j = a_j \text{for } 1 \leq j \leq i-1\big)$. Now use (1.29) to observe that

$$\big\|\widetilde{\mathscr{L}}(C_1, \ldots, C_{qT}) - \mathscr{L}(C'_1, \ldots, C'_{qT})\big\|_{TV} \leq \mathbb{P}\Big((C_1, \ldots, C_{qT}) \neq (C'_1, \ldots, C'_{qT})\Big)$$

$$= \sum_{i=1}^{qT} \mathbb{P}\Big(C_i \neq C'_i, (C_1, \ldots, C_{i-1}) = (C'_1, \ldots, C'_{i-1})\Big). \tag{1.30}$$

As noted above, either both C_i and C'_i are fresh or both are old. If both of them are old, then it has to be that on the event $(C_1, \ldots, C_{i-1}) = (C'_1, \ldots, C'_{i-1})$ the bits C_i and C'_i are equal. This is because on the preceding event, if $C_i = C(l,t)$ and $C'_i = C'(l,t)$ then $\mathrm{Al}_{t-1}(M_{l'}) = \mathrm{Th}_{t-1}(M_{l'})$ and $\mathrm{Al}_t(M_{l'}) = \mathrm{Th}_t(M_{l'})$ for every $l' < l$. So in this case the probability in the summand of (1.30) is zero. So assume that C_i and C'_i are both fresh bits. Use $\mathscr{C}_j, \mathscr{C}'_j$ and $\mathscr{a}_j$ to denote $(C_1, \ldots, C_j), (C_1, \ldots, C'_j)$ and $(a_1, \ldots, a_j)$ respectively. Then,

$$\begin{aligned}
&\mathbb{P}\Big(C_i \neq C'_i, \mathscr{C}_{i-1} = \mathscr{C}'_{i-1}\Big) \\
&= \sum_{\mathscr{a}_{i-1}} \mathbb{P}\big(C_i \neq C'_i \;\big|\; \mathscr{C}_{i-1} = \mathscr{C}'_{i-1} = \mathscr{a}_{i-1}\big)\mathbb{P}\big(\mathscr{C}_{i-1} = \mathscr{C}'_{i-1} = \mathscr{a}_{i-1}\big) \\
&\leq \sum_{\mathscr{a}_{i-1}} \mathbb{P}\big(C_i \neq C'_i \;\big|\; \mathscr{C}_{i-1} = \mathscr{C}'_{i-1} = \mathscr{a}_{i-1}\big)\mathbb{P}\big(\mathscr{C}'_{i-1} = \mathscr{a}_{i-1}\big) \\
&= \sum_{\mathscr{a}_{i-1}} \Big\| \mathscr{L}\big(C_i \mid \mathscr{C}_{i-1} = \mathscr{a}_{i-1}, \overline{\mathscr{P}}, \overline{\mathscr{S}}, L\big) - \mathscr{L}\big(C'_i \mid \mathscr{C}'_{i-1} = \mathscr{a}_{i-1}\big) \Big\|_{TV} \mathbb{P}\big(\mathscr{C}'_{i-1} = \mathscr{a}_{i-1}\big), \qquad (1.31)
\end{aligned}$$

where the last line follows from the fact that on the event $\mathscr{C}_{i-1} = \mathscr{C}'_{i-1}$, we coupled (C_i, C'_i) optimally. Since C'_i is a fresh bit, it is independent of $\mathscr{C}'_{i-1}$ and we can drop the $\mathscr{C}'_{i-1} = \mathscr{a}_{i-1}$ from $\mathscr{L}\big(C'_i \mid \mathscr{C}'_{i-1} = \mathscr{a}_{i-1}\big)$. Let $g(\mathscr{p}, S, z) = \mathbb{P}\big(\overline{\mathscr{P}} = \mathscr{p}, \overline{\mathscr{S}} = S, L = z\big)$. Then, continuing from above, we get

$$\begin{aligned}
(1.31) &= \mathbb{E}\left[\Big\| \mathscr{L}\big(C_i \mid \mathscr{C}_{i-1}, \overline{\mathscr{P}}, \overline{\mathscr{S}}, L\big) - \mathscr{L}\big(C'_i\big) \Big\|_{TV} \;\middle|\; \overline{\mathscr{P}}, \overline{\mathscr{S}}, L \right] \cdot g\big(\overline{\mathscr{P}}, \overline{\mathscr{S}}, L\big) \\
&\leq \mathbb{E}\left[\Big\| \mathscr{L}\big(C_i \mid \mathscr{C}_{i-1}, \overline{\mathscr{P}}, \overline{\mathscr{S}}, L\big) - \mathscr{L}\big(C'_i\big) \Big\|_{TV} \;\middle|\; \overline{\mathscr{P}}, \overline{\mathscr{S}}, L \right] \qquad (1.32)
\end{aligned}$$

Next, we will bound the expected value of the total variation distance above using Theorem 1.6.1. Let $\{C_{i_1}, C_{i_2}, \ldots, C_{i_o}\} \subseteq \{C_1, \ldots, C_{i-1}\}$ be a maximal set of fresh bits. That is any bit that is not in the above set is old. Let C_i and C'_i also be fresh. Then, from Theorem 1.6.1 we have the following inequality.

$$\mathbb{E}\Big\| \mathscr{L}\big(C_i \mid \mathscr{C}_{i-1}, \overline{\mathscr{P}}, \overline{\mathscr{S}}, L\big) - \mathscr{L}\big(C'_i\big) \Big\|_{TV}$$

$$
\begin{aligned}
&= \mathbb{E}\left\| \mathscr{L}\big(C_i \mid C_{i_1}, C_{i_2}, \ldots, C_{i_o}, \overline{\mathscr{P}}, \overline{\mathscr{S}}, L\big) - \mathscr{L}\big(C'_i\big) \right\| \\
&\leq 2 \cdot 2^{-\alpha(k,l+o,n)} \\
&\leq 2 \cdot 2^{-\alpha(k,l+i-1,n)}, \text{ since } o \leq i-1. \qquad (1.33)
\end{aligned}
$$

Therefore, by applying $\mathbb{E}$ to (1.32) we get

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{P}\left(C_i \neq C'_i, \mathscr{C}_{i-1} = \mathscr{C}'_{i-1} \right) \right] \\
&\leq \mathbb{E}\left\{ \mathbb{E}\left[\left\| \mathscr{L}\big(C_i \mid \mathscr{C}_{i-1}, \overline{\mathscr{P}}, \overline{\mathscr{S}}, L\big) - \mathscr{L}\big(C'_i\big) \right\|_{TV} \;\middle|\; \overline{\mathscr{P}}, \overline{\mathscr{S}}, L \right] \right\} \\
&\leq 2 \cdot 2^{-\alpha(k,l+i-1,n)}, \text{ from (1.33).} \qquad (1.34)
\end{aligned}
$$

Note that in the above string of inequalities, the probability distributions themselves depend on the values of $L, \overline{\mathscr{P}}, \overline{\mathscr{S}}$ since our coupling depended on it. We would like to clarify that we are not just taking the expectation of a constant. If we now apply expectations on (1.30) and use (1.34) above, we get

$$
\begin{aligned}
&\mathbb{E}\big\| \widetilde{\mathscr{L}}(C_1, \ldots, C_{qT}) - \mathscr{L}(C'_1, \ldots, C'_{qT}) \big\|_{TV} \\
&\leq \sum_{i=1}^{qT} 2 \cdot 2^{-\alpha(k,l+i-1,n)} \\
&\leq qT \cdot 2 \cdot 2^{-\alpha(k,l+qT-1,n)}.
\end{aligned}
$$

Since, $\|\mathscr{L} - \mathscr{L}'\| \leq \|\widetilde{\mathscr{L}}(C_1, \ldots, C_{qT}) - \mathscr{L}(C'_1, \ldots, C'_{qT})\|$ as noted in the beginning of this proof, our proof is complete. □

Our aim is to compare the expected total variation distance between the result of applying our algorithm to q messages and the result of applying a uniformly random permutation to the same. We have already bounded the distance between the Thorp shuffle and our algorithm. So by triangle inequality, all we need is a bound on the distance between the Throp shuffle and a uniform permutation. In [**25**], Morris, Rogaway and Stegers prove the following.

THEOREM 1.7.1. *[25] Let $\pi : \mathscr{M} \to \mathscr{M}$ be a uniformly random permutation on the set of messages $\mathscr{M}$ and let $\mathscr{U}$ be the distribution of $\big(\pi(M_1), \pi(M_2), \ldots, \pi(M_q)\big)$. Also let $\mathscr{L}'$ be the distribution of $\mathrm{Th}_T(M_1, \ldots, M_q)$ as in Lemma 1.7.1 above. Let $T = s(2m-1)$ for some whole number s, where $2^m = |\mathscr{M}|$. Then,*

$$\|\mathscr{L}' - \mathscr{U}\|_{TV} \leq \frac{q}{s+1}\left(\frac{4mq}{2^m}\right)^s.$$

The theorem above is closely related to the notion of mixing of Markov chains. In fact, the inequality in the theorem above gives a rate of mixing for the projected Thorp shuffle. For further works on mixing of the Thorp shuffle, refer to [**21**, **22**, **23**]. As discussed above, we now bound the expected total variation distance between our algorithm and a uniformly random permutation.

THEOREM 1.7.2. *Let $L, \overline{\mathscr{S}}, \overline{\mathscr{P}}$ be as in Lemma* 1.7.1. *Let $\mathscr{U}$ be the distribution of $\big(\pi(M_1), \ldots, \pi(M_q)\big)$ for a uniformly random permutation $\pi : \mathscr{M} \to \mathscr{M}$ and let $\mathscr{L}$ be the distribution of $\mathrm{Al}_T(M_1, \ldots, M_q)$ conditioned on $L, \overline{\mathscr{P}}, \overline{\mathscr{S}}$. Also let*

$$\alpha(k, a, n) = \frac{1}{8 \ln 2}\left(1 - \frac{a + 1 + n/(2 \ln 2)}{k - n}\right)^2,$$

where l is the length of L and n is the number of probes used to generate a single random bit in our algorithm. Assume that $l + qT + n/(2 \ln 2) \leq k - n$. Then for $T = s(2m-1)$, where s is a whole number, we have

$$\mathbb{E}\|\mathscr{L} - \mathscr{U}\|_{TV} \leq 2qT \cdot 2^{-n \cdot \alpha(k, l+qT-1, n)} + \frac{q}{s+1}\left(\frac{4mq}{2^m}\right)^s.$$

PROOF. This follows from Lemma 1.7.1 and Theorem 1.7.1 using triangle inequality for total variation distance.

$$\begin{aligned}\mathbb{E}\|\mathscr{L} - \mathscr{U}\|_{TV} &\leq \mathbb{E}\Big[\|\mathscr{L} - \mathscr{L}'\|_{TV} + \|\mathscr{L}' - \mathscr{U}\|_{TV}\Big] = \mathbb{E}\Big[\|\mathscr{L} - \mathscr{L}'\|_{TV}\Big] + \|\mathscr{L}' - \mathscr{U}\|_{TV} \\ &\leq 2qT \cdot 2^{-n \cdot \alpha(k, l+qT-1, n)} + \frac{q}{r+1}\left(\frac{4mq}{2^m}\right)^r.\end{aligned}$$

□

1.8. The Bound on Security

In this section we show that the notion of security we defined in Section 1.4 is the same as the expected total variation distance and hence as a corollary we have a bound on the security. We then prove the main theorem of this chapter.

LEMMA 1.8.1. *Let $\mathscr{A}$ be any adversary. As defined in section 1.4, let $\mathscr{A}(1)$ be the answer given by the adversary in world 1 when asked what what world he is in. Similarly let $\mathscr{A}(0)$ be the answer he gives while in world 0. Let $\mathscr{L}$ and $\mathscr{U}$ be as in Theorem 1.7.2, then*

$$\boldsymbol{Adv}(\mathscr{A}) = \mathbb{P}_1\big(\mathscr{A}(1) = 1\big) - \mathbb{P}\big(\mathscr{A}(0) = 1\big) \leq \mathbb{E}\|\mathscr{L} - \mathscr{U}\|.$$

PROOF. The adversary has access to the outputs, the random oracle and the leakage in each world. Having access to the random oracle means that potentially the adversary has access to all the probes and sub probe indices. If we condition on the probes $\overline{\mathscr{P}}$, the sub probe indices $\overline{\mathscr{S}}$ and the leakage $L = \Phi(K)$, then the algorithm the adversary uses to determine whether he is in world 0 or world 1 is a function of the outputs $(C_1, \ldots, C_q)$ that are provided to him in either of the worlds. Let $\mathscr{O}$ be the set of q-tuples of ciphertexts for which the adversary's answer is one. Then,

$$\begin{aligned}\mathbf{Adv}(\mathscr{A}) &= \mathbb{E}\Big[\mathbb{P}_1\big(\mathscr{A}(1) = 1 \mid \overline{\mathscr{P}}, \overline{\mathscr{S}}, L\big) - \mathbb{P}_0\big(\mathscr{A}(0) = 1\big)\Big] \\ &= \mathbb{E}\Big[\mathbb{P}_1\big((C_1, \ldots, C_q) \in \mathscr{O} \mid \overline{\mathscr{P}}, \overline{\mathscr{S}}, L\big) - \mathbb{P}_0\big((C_1, \ldots, C_q) \in \mathscr{O}\big)\Big].\end{aligned}$$

An alternate equivalent definition for total variation distance is as follows. For probability distributions μ, ν on a finite set Ω,

$$\|\mu - \nu\|_{TV} = \max_{A \subseteq \Omega} \big[\mu(A) - \nu(A)\big].$$

One can find the equivalence of the above with the definition we gave, in Section 4.1 of [**19**]. Combining the two equations above we get,

$$\begin{aligned}\mathbf{Adv}(\mathscr{A}) &= \mathbb{E}\Big[\mathbb{P}_1\big((C_1, \ldots, C_q) \in \mathscr{O} \mid \overline{\mathscr{P}}, \overline{\mathscr{S}}, L\big) - \mathbb{P}_0\big((C_1, \ldots, C_q) \in \mathscr{O}\big)\Big] \\ &\leq \mathbb{E}\Big\|\mathscr{L}(\mathrm{Al}_T(M_1), \ldots, \mathrm{Al}_T(M_q) \mid \overline{\mathscr{P}}, \overline{\mathscr{S}}, L) - \mathscr{L}(\pi(M_1), \ldots, \pi(M_q))\Big\| \\ &= \mathbb{E}\|\mathscr{L} - \mathscr{U}\|.\end{aligned}$$

□

So far, all the analysis we've done with total variation distance is under the assumption that the leakage function doesn't depend on the random oracle used to generate the probes/sub probes we are interested in. That is we assumed that the results of the random oracle calls we made in our algorithm were unknown to the adversary. In other words, we assumed that the probes involved in our algorithm were truly random. We will now prove the main result in which the adversary is allowed to make random oracle calls while choosing a leakage function.

PROOF. (Theorem 1.4.1) Assume that the adversary $\mathscr{A}$ can make at most r random oracle calls while deciding on a leakage function. Call a message M bad if the adversary makes a random oracle call with input $(b_2, b_3, \ldots, b_m, t)$ for some t and some message M such that $\mathrm{Al}_t(M) = (b_1, b_2, \ldots, b_n)$. Since the cipher is invertible for each time t, exactly two bad messages corresponding to each $(b_1, b_2, \ldots, b_{n-1}, t)$. This is because $\mathrm{Al}_t(M)$ can be either $(0, b_2, \ldots, b_m)$ or $(1, b_2, \ldots, b_m)$. Let $\mathscr{M}_b$ the set of bad messages corresponding to an adversary. Then, $|\mathscr{M}_b| \leq 2r$. Since the messages $M_1, \ldots, M_q$ are distinct and uniformly random, we get

$$\mathbb{P}\big(\, M_i \in \mathscr{M}_b \text{ for some } i = 1, 2, \ldots, q \,\big) = B(|\mathscr{M}|, q, m) \leq B(2r, q, m),$$

where B is as defined in (1.3). Let G be the event that $M_i \notin \mathscr{M}_b$ for $i = 1, 2, \ldots, q$. On the event G, the bound on total variation we obtained in Theorem 1.7.2 holds. Note that knowing a random oracle call that is not associated with any of $M_1, \ldots, M_q$ doesn't affect all our analysis till Theorem 1.7.2 since random oracles produce independent outputs when applied on different inputs. So, by Lemma 1.8.1 we get that for any fixed adversary $\mathscr{A}$

$$\begin{aligned}\mathbb{P}_1\big(\mathscr{A}(1) = 1 \,\big|G\,\big) - \mathbb{P}_0\big(\mathscr{A}(0) = 1 \,\big|\, G\big) &\leq \mathbb{E}\big[\|\mathscr{L} - \mathscr{U}\|\big] \\ &\leq 2qT \cdot 2^{-n\cdot\alpha(k, l+qT-1, n)} + \frac{q}{s+1}\left(\frac{4mq}{2^m}\right)^s.\end{aligned}$$

Putting all of this together, we have that for any adversary $\mathscr{A}$ who makes at most r random oracle calls,

$$\mathbb{P}\big(\mathscr{A}(1) = 1\big) - \mathbb{P}\big(\mathscr{A}(0) = 1\big) = \Big[\mathbb{P}\big(\mathscr{A}(1) = 1 \mid G\big) - \mathbb{P}\big(\mathscr{A}(0) = 1 \mid G\big)\Big]\mathbb{P}(G)$$

$$
\begin{aligned}
&+\Big[\mathbb{P}\big(\mathscr{A}(1)=1 \mid G^c\big)-\mathbb{P}\big(\mathscr{A}(0)=1 \mid G^c\big)\Big]\mathbb{P}(G^c)\\
\leq&\Big[\mathbb{P}\big(\mathscr{A}(1)=1 \mid G\big)-\mathbb{P}\big(\mathscr{A}(0)=1 \mid G\big)\Big]+\mathbb{P}(G^c)\\
\leq&2qT\cdot 2^{-n\cdot\alpha(k,l+qT-1,n)}+\frac{q}{s+1}\left(\frac{4mq}{2^m}\right)^s+\mathbb{P}(G^c)\\
\leq&2qT\cdot 2^{-n\cdot\alpha(k,l+qT-1,n)}+\frac{q}{s+1}\left(\frac{4mq}{2^m}\right)^s+B(2r,q,m).
\end{aligned}
$$

This is true for any adversary who makes at most r random oracle calls. So this also holds after one maximizes over all such adversaries. This completes the proof. □

CHAPTER 2

Transience of Simple Random Walks With Linear Entropy Growth

2.1. Introduction

In this chapter we show that if the entropy of the n^{th} step of a simple random walk on a connected graph with bounded degree is at least Cn for every n and some $C > 0$ which is independent of the starting position, then the walk is transient. This was conjectured by Benjamini Itai (personal communication, 2015). The entropy growth of Markov chains has been extensively studied for transitive chains on countable state space in works such as [**3**, **4**, **5**, **13**, **16**, **17**, **18**, **27**]. In fact, it is already known that for irreducible transitive Markov chains, linear entropy growth implies transience. Let $Z_0, Z_1, \ldots$ be an irreducible, transitive Markov chain starting at some state $Z_0 = z_0$ such that the entropy of Z_1 is finite. Then,

$$\hbar = \lim_{n\to\infty} \frac{\text{Entropy of } Z_n}{n}$$

exists and doesn't depend on z_0. $\hbar$ is known as Avez entropy in the literature. It first appeared in [**3**]. If $\hbar$ exists, then $\hbar > 0$ is equivalent to inequality (2.1) which forms the hypothesis of the main theorem of this chapter. Many important results about Avez entropy have been collected in the book [**20**]. One of the equivalences in Theorem 14.20 from [**20**] is that for the chain described above, $\hbar = 0$ if and only if the Liouville property holds. We say that a Markov chain has the Liouville property if all bounded harmonic functions on the state space of this Markov chain are constants. So, $\hbar > 0$ implies that there is a bounded non-constant harmonic function F on the state space of $(Z_i)_{i\geq 0}$. If the chain $(Z_i)_{i\geq 0}$ is recurrent, then by applying the Martingale convergence theorem on $\big(F(Z_i)\big)_{i\geq 0}$ one can conclude that F is a constant. Hence, we get that $(Z_i)_{i\geq 0}$ is transient. For random walk on groups, it was shown that $\hbar = 0$ implies the Liouville property for the first time in [**4**]. The equivalence of these two properties for random walks on groups was first shown in [**27**] and then later independently in [**13**]. The equivalence was then generalized to

transitive chains in [**17**]. For random walks on groups, the Avez entropy is also related to other important notions like drift/speed, spectral radius and volume growth. For sharp inequalities relating these quantities to each other, refer to [**16**].

Most of the work concerning relations between entropy growth and other interesting properties of a Markov chain seems to be in the setting of either transitive Markov chains or random walk on groups. Theorem 2.1.1 given below, which is the main theorem of this chapter, relates the entropy growth of a simple random walk on a connected graph to transience.

THEOREM 2.1.1. *Let $X_0, X_1, X_2, \ldots$ be the simple random walk on an infinite connected graph $G = (V, E)$ with maximum degree d, such that $X_0 = x_0$. Let E_n be the entropy of X_n, i.e., $E_n = \sum_{x \in V} -\mathbb{P}(X_n = x) \log\big(\mathbb{P}(X_n = x)\big)$. If*

$$E_n \geq Cn \tag{2.1}$$

for some C independent of x_0, then the random walk is transient.

Note that the entropy defined in the above theorem is a finite sum since each vertex has finite degree and hence the support of X_n is finite. So the entropy makes perfect sense and there is no question of convergence. Also note that the converse of the above theorem is obviously false because the simple random walk on $\mathbb{Z}^3$ is transient however since the n^{th} step of such a walk is supported on a set of order n^3, the maximum entropy of the n^{th} step is of the order $\log(n)$. Henceforth whenever the inequality (2.1) holds for some C independent of the starting position, we will say that the *linear entropy condition* holds. In Section 2.6 we give an example to show that the theorem fails to hold if the C in (2.1) is not independent of the starting position of the random walk.

2.2. Entropy and the Probability of Escape in n Steps

Definitions. Let P be the transition matrix of the Markov chain $X_0, X_1, \ldots$ defined above. Use $P^n(x, y)$ to denote $\mathbb{P}\big(X_n = y \big| \ X_0 = x\big)$. For any set $V' \subset V$, use $P^n(x, V')$ and $P^n(V', x)$ to denote $\sum_{v \in V'} P^n(x, v)$ and $\sum_{v \in V'} P^n(v, x)$ respectively.

Our proof of Theorem 2.1.1 uses a set valued process called the evolving set process, which

was used in [**24**] to obtain bounds on mixing times of Markov chains in terms of isoperimetric inequalities. The notion of evolving sets is related to strong stationary duals introduced by Fill and Diaconis [**14**]. Before we introduce evolving sets, let's make the following observation. Recall that d is the maximum degree of G.

LEMMA 2.2.1. *Assume that the linear entropy growth condition holds and that $X_0 = x_0$. Then for any set $A \subseteq V$,*

$$P^n(x_0, A^c) \geq \frac{Cn - \log 2|A|}{n \log d}. \tag{2.2}$$

PROOF. For any set A, let $B = \text{support}(X_n) \cap A$ and let $B' = \text{support}(X_n) \setminus A$. Then, by the definition of entropy we have

$$\begin{aligned} E_n &= \sum_{v \in B} -P^n(x_0, v) \log\big(P^n(x_0, v)\big) + \sum_{v \in B'} -P^n(x_0, v) \log\big(P^n(x_0, v)\big) \\ &= P^n(x_0, B) \sum_{v \in B} -\frac{P^n(x_0, v)}{P^n(x_0, B)} \log\left(\frac{P^n(x_0, v)}{P^n(x_0, B)}\right) + P^n(x_0, B') \sum_{v \in B'} -\frac{P^n(x_0, v)}{P^n(x_0, B')} \log\left(\frac{P^n(x_0, v)}{P^n(x_0, B')}\right) \\ &\quad - P^n(x_0, B) \log\big(P^n(x_0, B)\big) - P^n(x_0, B') \log\big(P^n(x_0, B')\big). \end{aligned}$$

Now observe that if μ is any probability distribution supported on Ω and X is a random variable with distribution μ, then $\sum_{x \in \Omega} \mu(x) \log(1/\mu(x)) = \mathbb{E} \log(1/\mu(X)) \leq \log\big(\mathbb{E}(1/\mu(X))\big) = \log|\Omega|$, by the Jensen's inequality applied to $-\log(\cdot)$. This is in fact a standard result from information theory, which is stated as Theorem 2.6.4 in [**11**]. Applying the preceding inequality to $P^n(x_0, \cdot)/P^n(x_0, B)$ and $P^n(x_0, \cdot)/P^n(x_0, B')$, which are probability measures supported on B and B' respectively, we get

$$E_n \leq P^n(x_0, B) \log|B| + P^n(x_0, B') \log|B'| - P^n(x_0, B) \log\big(P^n(x_0, B)\big) - P^n(x_0, B') \log\big(P^n(x_0, B')\big).$$

Note that $P^n(x_0, B') = P^n(x_0, A^c)$. Also observe that if we set $q = P^n(x_0, B)$, then

$$-P^n(x_0, B) \log\big(P^n(x_0, B)\big) - P^n(x_0, B') \log\big(P^n(x_0, B')\big) = -q \log(q) - (1-q) \log(1-q) \leq \log 2$$

which again follows from Jensen's inequality. Combining the two inequalities above, we get,

$$\begin{aligned} E_n &\leq \log|B| + P^n(x_0, A^c)\log|B'| + \log 2 \\ &\leq \log|A| + P^n(x_0, A^c)\log|\text{support}(X_n)| + \log 2. \end{aligned}$$

Since d is the maximum degree of each vertex, we have $|\text{support}(X_n)| \leq d^n$. Hence,

$$E_n \leq \log 2|A| + n\log(d)P^n(x_0, A^c).$$

Using the hypothesis that $E_n \geq Cn$, and rearranging, we get the desired result. □

COROLLARY 2.2.1. *Assume that the linear entropy growth condition holds. This time, also assume that X_0 is random. Then for any set $A \subseteq V$,*

$$\mathbb{P}(X_n \in A^c) \geq \frac{Cn - \log 2|A|}{n\log d}. \tag{2.3}$$

PROOF. This follows by conditioning on $\{X_0 = x_0\}$ and using the fact that the right side of (2.2) is independent of x_0. □

2.3. The Evolving Set Process

Let $U_1, U_2, \ldots$ be i.i.d Uniform($[0,1]$) random variables. For any stationary measure $\pi : V \to V$, define $Q_1(x,y) = Q(x,y) = \pi(x)P(x,y)$ and $Q_t(x,y) = \pi(x)P^t(x,y)$. For any set $B \subset V$, use $Q_t(B,y)$ to denote $\sum_{x\in B} Q_t(x,y)$. Define $Q(y,B)$ analogously.

Definition. (Evolving Sets) For a fixed set S we now define an integer valued process $T_0, T_1, \ldots$ and a set valued process $S_{T_0}, S_{T_1}, \ldots$ inductively as follows. Set $T_0 = 0$ and $S_{T_0} = S$. Now assuming T_{m-1} and $S_{T_{m-1}}$ are given, we define $T_m = T_{m-1} + L_m$, where $L_1, L_2, \ldots$ is an integer valued process such that L_j is a function of $S_{T_{j-1}}$ and $L_j \geq 0$ for every $j \geq 1$. We then define S_{T_m} by imposing the condition that $y \in S_{T_m}$ if and only if $Q_{L_m}(S_{T_{m-1}}, y) \geq U_m\pi(y)$. The construction used here is similar to the one in [**24**], with the only difference being that we use a different transition matrix at each step of the evolving set process.

Note that the definition of the set valued process depends on the sequence of random variables $L_1, L_2, \ldots$. The results in this section are true for any such sequence of random variables as long

as they satisfy the minor conditions imposed in the definition above. From Section 2.4 onwards, we choose a particular sequence $\big(L_j\big)_{j\geq 1}$. For every integer $t \geq 0$, define

$$a(t) = \max\{i \ : \ T_i \leq t\}.$$

Then we have the following lemma.

LEMMA 2.3.1. *With the notation as above, the following is true:*

$$Q_t(S, y) = \mathbb{E}\big[Q_{t-T_{a(t)}}(S_{T_{a(t)}}, y)\big]. \tag{2.4}$$

PROOF. We will prove this using strong induction. That is, we will assume that the statement holds for $t \leq n$ and then prove it for $t = n+1$. Let M be the first time $L_m \neq 0$, i.e., $L_1 = L_2 = \cdots = L_{M-1} = 0$ but $L_M \neq 0$. We will take $t \leq T_M$ to be the base case. If $L_m = 0$ then $S_{T_m} = S_{T_{m-1}}$ and hence $S_{T_0} = S_{T_1} = \cdots = S_{T_{M-1}} = S$. Note that T_M is not random since T_M is a function of $S_{T_{M-1}} = S$ which is deterministic. For $t < T_M$, note that $a(t) = M - 1$. So, $T_{a(t)} = T_{M-1} = 0$ and $S_{T_{a_t}} = S_0 = S$. Hence, in this case the statement is trivial. Now for $t = T_M$, note that $a(t) = M$, $L_M = T_M$ and

$$\begin{aligned}
\mathbb{E}\big[Q_{t-T_M}(S_{T_M}, y)\big] = \mathbb{E}\big[Q_0(S_{T_M}, y) &= \mathbb{E}\Big[\sum_{z\in S_{T_M}} \pi(z)\mathbb{1}_{\{z=y\}}\Big] \\
&= \mathbb{E}\big[\pi(y)\mathbb{1}_{y\in S_{T_M}}\big] = \pi(y)\mathbb{P}(y \in S_{T_M}) \\
&= \pi(y)\frac{Q_{L_M}(S_{T_{M-1}}, y)}{\pi(y)} = Q_{T_M}(S, y) = Q_t(S, y).
\end{aligned}$$

Now assume that the statement holds any $t \leq n$ and for any process $S_{T_1}, S_{T_2}, \ldots$ which satisfies the above conditions. Since $t \leq T_M$ is the base case, we can assume $n \geq T_M$. Now condition on S_{T_M} and define the new process $\widetilde{S}_{\widetilde{T}_0}, \widetilde{S}_{\widetilde{T}_1}, \ldots$ by setting $\widetilde{S}_{\widetilde{T}_m} = S_{T_{M+m}}$, $\widetilde{T}_m = T_{M+m} - T_M$ and $\widetilde{L}_m = L_{M+m}$. If we now define

$$\widetilde{a}(t) = \max\{i \ : \ \widetilde{T}_i \leq t\},$$

then, $\widetilde{a}(t) = a(t + T_M) - M$ and $\widetilde{T}_{\widetilde{a}(t)} = T_{a(t+T_M)} - T_M$. Let $n + 1 = T_M + t$, then $t \leq n$ since $T_M \geq 1$. Moreover the process defined above satisfies the hypothesis and hence by induction, we

have

$$Q_t(S_{T_M}, y) = \mathbb{E}\Big[Q_{t-\widetilde{T}_{\widetilde{a}(t)}}(\widetilde{S}_{\widetilde{T}_{\widetilde{a}(t)}}, y) \mid S_{T_M}\Big] = \mathbb{E}\Big[Q_{t+T_M-T_{a(t+T_M)}}(S_{T_{a(t+T_M)}}, y) \mid S_{T_M}\Big]. \tag{2.5}$$

Next, note that

$$\begin{aligned} Q_{n+1}(S,y) = Q_{T_M+t}(S,y) &= \sum_{z\in\Omega} Q_{T_M}(S,z)P^t(z,y) \\ &= \sum_{z\in\Omega} Q_t(z,y)\frac{Q_{T_M}(S,z)}{\pi(z)} = \sum_{z\in\Omega} Q_t(z,y)\frac{Q_{L_M}(S,z)}{\pi(z)} = \sum_{z\in\Omega} Q_t(z,y)\mathbb{P}(z\in S_{T_M}) \\ &= \mathbb{E}\Big[\sum_{z\in\Omega} Q_t(z,y)\mathbb{1}_{\{z\in S_{T_M}\}}\Big] = \mathbb{E}\big[Q_t(S_{T_M},y)\big]. \end{aligned} \tag{2.6}$$

Applying $\mathbb{E}$ to equation (2.5) and continuing from (2.6) we get

$$\begin{aligned} Q_{n+1}(S,y) &= \mathbb{E}\Big[Q_{t+T_M-T_{a(t+T_M)}}(S_{T_{a(t+T_M)}}, y)\Big] \\ &= \mathbb{E}\big[Q_{n+1-T_{a(n+1)}}(S_{T_{a(n+1)}}, y)\big], \end{aligned}$$

which completes the induction. □

Corollary 2.3.1. *Let $S_{T_0}, S_{T_1}, \ldots$ and $a(\cdot)$ be as in lemma 2.3.1 above. Assume this time that $S_{T_0} = S = \{x_0\}$. Then,*

$$P^t(x_0, y) = \frac{1}{\pi(x_0)}\mathbb{E}\big[Q_{t-T_{a(t)}}(S_{T_{a(t)}}, y)\big].$$

Proof. Let $S = \{x_0\}$ in lemma 2.3.1 above. □

2.4. Relating Transience and Evolving Sets

Corollary 2.3.1 is useful since it will help us bound $\sum_{t\geq 0} P^t(z,y)$, the existence of which implies transience. In the Lemma below we relate this sum to $\sum_{i=0}^{\infty}\mathbb{E}\big[\sqrt{\pi(S_{T_i})}\big]$, for a specific choice of π and $(L_j)_{j\geq 1}$. Henceforth we will fix π and $(L_j)_{j\geq 1}$ to be the following :

$$\pi(x) = \text{degree}(x) \text{ and}$$

$$L_m = 2\lceil \log(2\cdot\pi(S_{T_{m-1}}))/C\rceil \text{ for } m \geq 1.$$

LEMMA 2.4.1. *Let $S_{T_0}, S_{T_1}, \ldots$ be as above and let C be as in theorem 2.1.1. Then for any $y \in \Omega$,*

$$\sum_{t=0}^{\infty} P^t(x_0, y) \leq 4d \left\lceil \frac{1}{C} \right\rceil \sum_{i=0}^{\infty} \mathbb{E}\left[\sqrt{\pi(S_{T_i})}\right]. \tag{2.7}$$

PROOF. Note that for any $m \geq 0$, $S' \subset V$ and $y \in V$, we have $Q_m(S', y) \leq Q_m(V, y) = \pi(x)$. The equality in the preceding line is due to the fact that π is a stationary measure. Moreover, $Q_m(\emptyset, y) = 0$. Therefore, we have

$$Q_{t-T_{a(t)}}(S_{T_{a(t)}}, y) \leq Q_{t-T_{a(t)}}(V, y) \mathbb{1}_{\left\{S_{T_{a(t)}} \neq \emptyset\right\}} = \pi(y) \mathbb{1}_{\left\{S_{T_{a(t)}} \neq \emptyset\right\}}. \tag{2.8}$$

Using (2.8) above and Corollary 2.3.1, we get

$$\begin{aligned} \sum_{t=0}^{\infty} P^t(x_0, y) &= \frac{1}{\pi(x_0)} \sum_{t=0}^{\infty} \mathbb{E}\left[Q_{t-T_{a(t)}}(S_{T_{a(t)}}, y)\right] \\ &\leq \frac{\pi(y)}{\pi(x_0)} \sum_{t=0}^{\infty} \mathbb{E}\left[\mathbb{1}_{\left\{S_{T_{a(t)}} \neq \emptyset\right\}}\right] \\ &\leq d \sum_{t=0}^{\infty} \mathbb{E}\left[\mathbb{1}_{\left\{S_{T_{a(t)}} \neq \emptyset\right\}}\right], \end{aligned} \tag{2.9}$$

where the last inequality follows from the fact that $1 \leq \pi(z) \leq d$ for any z. Next, observe that when $S_{T_m} \leq t < S_{T_{m+1}}$, by definition $a(t) = m$ and hence

$$\sum_{t=0}^{\infty} \mathbb{1}_{\left\{S_{T_{a(t)}} \neq \emptyset\right\}} = \sum_{i=0}^{\infty} (T_{i+1} - T_i) \mathbb{1}_{\{S_{T_i} \neq \emptyset\}} = \sum_{i=0}^{\infty} L_i \mathbb{1}_{\{S_{T_i} \neq \emptyset\}}.$$

So, continuing from (2.9) we get

$$\sum_{t=0}^{\infty} P^t(x_0, y) \leq d \cdot \mathbb{E}\left[\sum_{i=0}^{\infty} L_i \mathbb{1}_{\{S_{T_i} \neq \emptyset\}}\right] = d \cdot \mathbb{E}\left[\sum_{i=0}^{\infty} 2 \left\lceil \frac{\log(2 \cdot \pi(S_{T_i}))}{C} \right\rceil \mathbb{1}_{\{S_{T_i} \neq \emptyset\}}\right]. \tag{2.10}$$

Now observe the following fact about the natural logarithm which will be useful for bounding the above.

$$2\sqrt{x} \geq \lceil \log(2x) \rceil \quad \text{for } x \geq 1. \tag{2.11}$$

To prove this, observe first that $2\sqrt{x} \geq \log(2x) + 1$ for $x \geq 1$. This is true since the inequality holds for $x = 1$ and $d/dx(2\sqrt{x} - \log(2x) - 1) = \sqrt{2}/\sqrt{x} - 1/x \geq 0$ when $x \geq 1$. Since $\lceil \log(2x) \rceil \leq \log(2x) + 1$, the previous inequality gives us (2.11).

We can finally use (2.10) and (2.11) above to get

$$\sum_{t=0}^{\infty} P^t(x_0, y) \leq d \cdot \mathbb{E}\Big[\sum_{i=0}^{\infty} 2\left\lceil \frac{\log(2 \cdot \pi(S_{T_i}))}{C} \right\rceil \mathbb{1}_{\{S_{T_i} \neq \emptyset\}}\Big] \leq d \cdot \mathbb{E}\left[\sum_{i=0}^{\infty} 2\left\lceil \frac{1}{C} \right\rceil \lceil \log(2 \cdot \pi(S_{T_i})) \rceil \mathbb{1}_{\{S_{T_i} \neq \emptyset\}}\right]$$
$$\leq 4d\left\lceil \frac{1}{C} \right\rceil \mathbb{E}\Big[\sum_{i=0}^{\infty} \sqrt{\pi(S_{T_i})}\mathbb{1}_{\{S_{T_i} \neq \emptyset\}}\Big] = 4d\left\lceil \frac{1}{C} \right\rceil \sum_{i=0}^{\infty} \mathbb{E}\Big[\sqrt{\pi(S_{T_i})}\Big].$$

□

2.5. Decay of $\mathbb{E}\big[\sqrt{\pi(S_{T_i})}\big]$

We will show that $\mathbb{E}\big[\sqrt{\pi(S_{T_i})}\big]$ decays exponentially in i, by proving the following theorem.

THEOREM 2.5.1. *Let π, d, C, S_{T_j} be as above. Then, there exists a constant $0 \leq \alpha < 1$ depending only on C and d such that*

$$\mathbb{E}\Big[\sqrt{\pi(S_{T_m})} \;\Big|\; S_{T_{m-1}}\Big] \leq \alpha \cdot \sqrt{\pi(S_{T_{m-1}})}. \tag{2.12}$$

We will come back to the proof of theorem 2.5.1 after we prove a lemma. But first, let's set up some notation for this section. Throughout this section let

$$\widetilde{S} = S_{T_m},\ S = S_{T_{m-1}} \text{ and } \widetilde{\mathbb{E}}[\,\cdot\,] = \mathbb{E}[\,\cdot \mid S]. \tag{2.13}$$

Then,

LEMMA 2.5.1.

$$\widetilde{\mathbb{E}}\left[\left(\frac{\pi(\widetilde{S})}{\pi(S)} - 1\right)^{+}\right] \geq \frac{C}{2\log(d)} \tag{2.14}$$

for every S.

PROOF. Let $\widetilde{X}_0, \widetilde{X}_1, \ldots$ be a Markov chain with P as the transition matrix and $\mathbb{P}(\widetilde{X}_0 = y) = [\pi(y)/\pi(S)]\mathbb{1}_{\{y \in S\}}$. Let $s_1, s_2, \ldots$ be an enumeration of the elements of our graph in decreasing

order of $Q_{L_m}(S, s_i)/\pi(s_i)$. Let U be the uniform[0,1] random variable used to generate $\widetilde{S}$ from S. Define τ to be the minimum integer such that $\sum_{i=1}^{\tau} \pi(s_i) \geq \pi(S)$. That is,

$$\tau = \min\left\{k \ : \ \sum_{i=1}^{k} \pi(s_i) \geq \pi(S)\right\}.$$

Define

$$v = Q_{L_m}(S, s_\tau)/\pi(s_\tau).$$

By the definition of $\widetilde{S}$, $s_i \in \widetilde{S}$ whenever $U \leq Q_{L_m}(S, s_i)/\pi(s_i)$. Note that $Q_{L_m}(S, s_i)/\pi(s_i) \geq Q_{L_m}(S, s_\tau)/\pi(s_\tau)$ for every $i \leq \tau$. So, we can conclude that $s_1, \ldots, s_\tau \in \widetilde{S}$ whenever $U \leq v$. In this case $\pi(\widetilde{S}) \geq \sum_{i=1}^{\tau} \pi(s_i) \geq \pi(S)$. If $U > v$, then $\widetilde{S} = \{s_1, \ldots, s_k\}$ for $k < \tau$ and hence by the definition of τ, $\pi(\widetilde{S}) = \sum_{i=1}^{k} \pi(s_i) < \pi(S)$. So we have

$$\mathbb{1}_{\{\pi(\widetilde{S}) \geq \pi(S)\}} = \mathbb{1}_{\{U \leq v\}} \tag{2.15}$$

and

$$\begin{aligned}(\pi(\widetilde{S}) - \pi(S))^+ &= (\pi(\widetilde{S}) - \pi(S))\mathbb{1}_{\{\pi(\widetilde{S}) \geq \pi(S)\}} = (\pi(\widetilde{S}) - \pi(S))\mathbb{1}_{\{U \leq v\}} \\ &= \mathbb{1}_{\{U \leq v\}} \sum_{i=1}^{\tau} \mathbb{1}_{\{s_i \in \widetilde{S}\}} \pi(s_i) + \mathbb{1}_{\{U \leq v\}} \sum_{i > \tau} \mathbb{1}_{\{s_i \in \widetilde{S}\}} \pi(s_i) - \mathbb{1}_{\{U \leq v\}} \pi(S). \end{aligned} \tag{2.16}$$

Note that as we remarked above, if $i \leq \tau$ then $s_i \in \widetilde{S}$ whenever $U \leq v$. Therefore for $i \leq \tau$, we have $\{s_i \in \widetilde{S}, U \leq v\} = \{U \leq v\}$. This allows us to simplify the product of indicators in the first term above and get

$$\begin{aligned}(2.16) &= \mathbb{1}_{\{U \leq v\}} \left[\sum_{i=1}^{\tau} \pi(s_i) - \pi(S)\right] + \mathbb{1}_{\{U \leq v\}} \sum_{i > \tau} \mathbb{1}_{\{s_i \in \widetilde{S}\}} \pi(s_i) \geq \mathbb{1}_{\{U \leq v\}} \sum_{i > \tau} \mathbb{1}_{\{s_i \in \widetilde{S}\}} \pi(s_i) \\ &= \sum_{i > \tau} \mathbb{1}_{\{s_i \in \widetilde{S},\ U \leq v\}} \pi(s_i).\end{aligned}$$

For $i > \tau$, $Q_{L_m}(S, s_i)/\pi(s_i) \leq v$. So, $s_i \in \widetilde{S} \implies U \leq Q_{L_m}(S, s_i)/\pi(s_i) \implies U \leq v$. Hence, $\{s_i \in \widetilde{S},\ U \leq v\} = \{s_i \in \widetilde{S}\}$ and continuing from above we get

$$\widetilde{\mathbb{E}}\big[(\pi(\widetilde{S}) - \pi(S))^+\big] \geq \widetilde{\mathbb{E}} \sum_{i > \tau} \mathbb{1}_{\{s_i \in \widetilde{S}\}} \pi(s_i) = \sum_{i > \tau} \mathbb{P}\big(s_i \in \widetilde{S} \mid S\big)\pi(s_i) = \sum_{i > \tau} \frac{Q_{L_m}(S, s_i)}{\pi(s_i)} \pi(s_i)$$

$$= \pi(S) \sum_{i>\tau} \frac{Q_{L_m}(S, s_i)}{\pi(S)}. \tag{2.17}$$

Now observe that

$$\frac{Q_{L_m}(S, s_i)}{\pi(S)} = \sum_{y \in S} \frac{\pi(y) P^{L_m}(y, s_i)}{\pi(S)} = \sum_y [\pi(y)/\pi(S)] \mathbb{1}_{\{y \in S\}} \cdot P^{L_m}(y, s_i)$$

$$= \sum_y \mathbb{P}(\widetilde{X}_0 = y) P^{L_m}(y, s_i) = \mathbb{P}(\widetilde{X}_{L_m} = s_i). \tag{2.18}$$

So, from (2.17) we get

$$\widetilde{\mathbb{E}}\big[(\pi(\widetilde{S}) - \pi(S))^+\big] \geq \pi(S) \sum_{i>\tau} \mathbb{P}\big(\widetilde{X}_{L_m} = s_i\big). \tag{2.19}$$

Let $S' = \{s_1, \ldots, s_{\pi(S)}\}$. Note that $\tau \leq \pi(S)$ since $\pi(s) \geq 1$ for every s. Hence, $\{s_1, \ldots, s_\tau\} \subset \{s_1, \ldots, s_{\pi(S)}\}$ and $\sum_{i>\tau} \mathbb{P}\big(\widetilde{X}_{L_m} = s_i\big) \geq \sum_{i>\pi(S)} \mathbb{P}\big(\widetilde{X}_{L_m} = s_i\big)$. If we apply corollary 2.2.1 with $A = S'$, we get

$$\begin{aligned}
\sum_{i>\pi(S)} \mathbb{P}\big(\widetilde{X}_{L_m} = s_i\big) = \mathbb{P}(\widetilde{X}_{L_m} \in (S')^c) &\geq \frac{CL_m - \log(2|S'|)}{L_m \log(d)} \\
&= \frac{C\big(2\lceil \log(2 \cdot \pi(S))/C \rceil\big) - \log(2 \cdot \pi(S))}{\big(2\lceil \log(2 \cdot \pi(S))/C \rceil\big) \log(d)} \\
&\geq \frac{C\big(2\lceil \log(2 \cdot \pi(S))/C \rceil\big) - \lceil \log(2 \cdot \pi(S)) \rceil}{\big(2\lceil \log(2 \cdot \pi(S))/C \rceil\big) \log(d)} \\
&\geq \frac{C\big(2\lceil \log(2 \cdot \pi(S))/C \rceil\big) - C\lceil \log(2 \cdot \pi(S))/C \rceil}{\big(2\lceil \log(2 \cdot \pi(S))/C \rceil\big) \log(d)} = \frac{C}{2\log(d)}.
\end{aligned} \tag{2.20}$$

Combining (2.19) and (2.20) we get

$$\widetilde{\mathbb{E}}\big[(\pi(\widetilde{S}) - \pi(S))^+\big] \geq \pi(S) \frac{C}{2\log(d)}.$$

Finally, since $\widetilde{\mathbb{E}}$ is the conditional expectation given S, we can rearrange the above to get what we want. □

Proof. (Theorem 2.5.1) Let $S, \widetilde{S}$ be as before. Also let

$$x = \widetilde{\mathbb{E}}\left[\left(\frac{\pi(\widetilde{S})}{\pi(S)} - 1\right)^+\right] \text{ and } p = \mathbb{P}\Big(\pi(\widetilde{S}) \geq \pi(S) \;\Big|S\Big).$$

Note that $p = \widetilde{\mathbb{E}}\big[\mathbb{1}_{\pi(\widetilde{S})\geq\pi(S)}\big]$ and

$$p + x = \widetilde{\mathbb{E}}\left[\left(\frac{\pi(\widetilde{S})}{\pi(S)} - 1\right)\mathbb{1}_{\{\pi(\widetilde{S})\geq\pi(S)\}}\right] + \widetilde{\mathbb{E}}\big[\mathbb{1}_{\{\pi(\widetilde{S})\geq\pi(S)\}}\big] = \widetilde{\mathbb{E}}\left[\left(\frac{\pi(\widetilde{S})}{\pi(S)}\right)\mathbb{1}_{\{\pi(\widetilde{S})\geq\pi(S)\}}\right]. \tag{2.21}$$

Next observe that

$$\widetilde{\mathbb{E}}[\pi(\widetilde{S})] = \widetilde{\mathbb{E}}\Big[\sum_{y\in V}\mathbb{1}_{\{y\in\widetilde{S}\}}\pi(y)\Big] = \sum_{y\in V}\mathbb{P}\big(y\in\widetilde{S}\mid S\big)\pi(y) = \sum_{y\in V}Q_{L_m}(S,y) = \pi(S)\sum_{y\in V}Q_{L_m}(S,y)/\pi(S).$$

If $\widetilde{X}_i$ is as in the proof of lemma 2.5.1 then from (2.18) we see that

$$\widetilde{\mathbb{E}}[\pi(\widetilde{S})] = \pi(S)\sum_{y\in V}\mathbb{P}(\widetilde{X}_{L_m} = y) = \pi(S).$$

This give us

$$\widetilde{\mathbb{E}}\left(\frac{\pi(\widetilde{S})}{\pi(S)}\right) = 1. \tag{2.22}$$

Let $q = 1 - p$. Then,

$$\begin{aligned} q - x = 1 - (p + x) &= \widetilde{\mathbb{E}}\left(\frac{\pi(\widetilde{S})}{\pi(S)}\right) - \widetilde{\mathbb{E}}\left[\left(\frac{\pi(\widetilde{S})}{\pi(S)}\right)\mathbb{1}_{\{\pi(\widetilde{S})\geq\pi(S)\}}\right] \quad \text{from (2.22)} \\ &= \widetilde{\mathbb{E}}\left[\left(\frac{\pi(\widetilde{S})}{\pi(S)}\right)\mathbb{1}_{\{\pi(\widetilde{S})<\pi(S)\}}\right]. \end{aligned} \tag{2.23}$$

Note that $x \geq C/\big(2\log(d)\big)$ due to lemma 2.5.1. Also, $q - x \geq 0$ from (2.23). Hence, $C/\big(2\log(d)\big) \leq x \leq q$. Let $\widetilde{\mathbb{P}}(\cdot) = \mathbb{P}(\cdot \mid S)$, then we have

$$\begin{aligned} \widetilde{\mathbb{E}}\left[\sqrt{\pi(\widetilde{S})}\right] &= \widetilde{\mathbb{E}}\left[\sqrt{\pi(\widetilde{S})} \;\middle|\; \pi(\widetilde{S}) \geq \pi(S)\right]\widetilde{\mathbb{P}}\big(\pi(\widetilde{S}) \geq \pi(S)\big) + \widetilde{\mathbb{E}}\left[\sqrt{\pi(\widetilde{S})} \;\middle|\; \pi(\widetilde{S}) < \pi(S)\right]\widetilde{\mathbb{P}}\big(\pi(\widetilde{S}) < \pi(S)\big) \\ &= p\cdot\widetilde{\mathbb{E}}\left[\sqrt{\pi(\widetilde{S})} \;\middle|\; \pi(\widetilde{S}) \geq \pi(S)\right] + q\cdot\widetilde{\mathbb{E}}\left[\sqrt{\pi(\widetilde{S})} \;\middle|\; \pi(\widetilde{S}) < \pi(S)\right] \\ &\leq p\sqrt{\widetilde{\mathbb{E}}\big[\pi(\widetilde{S}) \mid \pi(\widetilde{S}) \geq \pi(S)\big]} + q\sqrt{\widetilde{\mathbb{E}}\big[\pi(\widetilde{S}) \mid \pi(\widetilde{S}) < \pi(S)\big]} \quad \text{by Jensen's inequality} \\ &= p\sqrt{\frac{\widetilde{\mathbb{E}}\big[\pi(\widetilde{S})\mathbb{1}_{\pi(\widetilde{S})\geq\pi(S)}\big]}{p}} + q\sqrt{\frac{\widetilde{\mathbb{E}}\big[\pi(\widetilde{S})\mathbb{1}_{\pi(\widetilde{S})<\pi(S)}\big]}{q}} \\ &= \sqrt{p}\sqrt{\pi(S)(p+x)} + \sqrt{q}\sqrt{\pi(S)(q-x)} \quad \text{from (2.21) and (2.23).} \end{aligned}$$

Applying Taylor's theorem to $\sqrt{1+z}$ about $z=0$, we get $0 < x_1/p < x/p$ and $0 < x_2/q < x/q$ such that

$$p\sqrt{1+\frac{x}{p}} = p\Big(1+\frac{x}{2p}-\frac{x^2}{8p^2}\big(1+x_1/p\big)^{-3/2}\Big) \text{ and}$$
$$q\sqrt{1-\frac{x}{q}} = q\Big(1-\frac{x}{2q}-\frac{x^2}{8q^2}\big(1-x_2/q\big)^{-3/2}\Big).$$

$(1+x_1/p)^{-3/2} \geq (1+q/p)^{-3/2} = p^{3/2}$ since $x_1 < x \leq q = 1-p$ and similarly $(1-x_2/q)^{-3/2} \geq 1$ since $x_2 > 0$. So, we get

$$p\sqrt{1+\frac{x}{p}} \leq p\Big(1+\frac{x}{2p}-\frac{x^2}{8p^2}p^{3/2}\Big) \text{ and}$$
$$q\sqrt{1-\frac{x}{q}} \leq q\Big(1-\frac{x}{2q}-\frac{x^2}{8q^2}\Big).$$

Adding the two inequalities above and multiplying by $\sqrt{\pi(S)}$, we get

$$\begin{aligned}
\widetilde{\mathbb{E}}\Big[\sqrt{\pi(\widetilde{S})}\Big] &= \sqrt{\pi(S)}\Big[p\sqrt{1+\frac{x}{p}}+q\sqrt{1-\frac{x}{q}}\Big] \\
&\leq \sqrt{\pi(S)}\Big[p+q-\frac{x^2}{8}\sqrt{p}-\frac{x^2}{8q}\Big] = \sqrt{\pi(S)}\Big[1-\frac{x^2}{8}\sqrt{p}-\frac{x^2}{8q}\Big] \\
&\leq \sqrt{\pi(S)}\big(1-x^2/8\big) \quad \text{since } \sqrt{p}\geq 0 \text{ and } 1/q \geq 1.
\end{aligned}$$

As discussed above, $x \geq C/(2\log(d))$ due to lemma 2.5.1. Hence,

$$\widetilde{\mathbb{E}}\Big[\sqrt{\pi(\widetilde{S})}\Big] \leq \sqrt{\pi(S)}\Big(1-C^2/32\log^2(d)\Big) = \alpha\sqrt{\pi(S)}. \tag{2.24}$$

□

Corollary 2.5.1. *Let π, d, C, S_{T_j} be as in section 2.3. Then,*

$$\mathbb{E}\Big[\sqrt{\pi(S_{T_m})}\Big] \leq \alpha^m \cdot \pi(x_0) \tag{2.25}$$

where $\alpha = 1 - C^2/32\log^2 d$ is a non-zero constant less than 1 and independent of the starting position.

PROOF. The proof is by induction. For $m = 0$, $S_{T_0} = S_0 = \{x_0\}$. So, $\mathbb{E}[\sqrt{\pi(S_{T_0})}] = \pi(x_0) \leq \alpha^0\pi(x_0)$. Now, if we assume the statement for $m-1$ then by the tower property of conditional expectation, we get

$$\begin{aligned}\mathbb{E}\big[\sqrt{\pi(S_{T_m})}\big] &= \mathbb{E}\Big[\mathbb{E}\big[\sqrt{\pi(S_{T_m})} \mid S_{T_{m-1}}\big]\Big] \leq \mathbb{E}\Big[\alpha \cdot \sqrt{\pi(S_{T_{m-1}})}\Big] \\ &\leq \alpha \cdot \alpha^{m-1}\pi(x_0) \quad \text{by induction.}\end{aligned}$$

□

The main theorem of this chapter can now be easily proved using lemma 2.4.1 and corollary 2.5.1 as follows.

PROOF. (Theorem 2.1.1) To show that x_0 is transient, it is sufficient to show that $\sum_{t=0}^{\infty} P^t(x_0, x_0) < \infty$. By Lemma 2.4.1 and Corollary 2.5.1 we get

$$\begin{aligned}\sum_{t=0}^{\infty} P^t(x_0, x_0) &\leq 4d\left\lceil \frac{1}{C} \right\rceil \sum_{i=0}^{\infty} \mathbb{E}\Big[\sqrt{\pi(S_{T_i})}\Big] \\ &\leq 4d\left\lceil \frac{1}{C} \right\rceil \sum_{i=0}^{\infty} \alpha^i\pi(x_0) < \infty\end{aligned}$$

since $0 \leq \alpha < 1$. Since the graph is connected, this means all the vertices are transient. Or alternatively just note that x_0 is arbitrary. □

2.6. Necessity of the Uniformity of C in the Linear Entropy Growth Condition

For each $n \geq 1$, let T_n be a full binary tree with height $B_n - 1$ and root r_n, where B_n is inductively defined by setting $B_1 = 1$ and

$$B_{n+1} = 32n^2 2^{B_n}.$$

Let $a_1, a_2, \ldots$ be a collection of vertices disjoint from the $T_n's$. Form a graph by first forming the edges $a_i a_{i+1}$ for every i and then the edges $a_j r_j$ for every j. In other words, we form a graph by taking the infinite path $a_1 a_2 a_3 \cdots$ and joining each a_j with the root of T_j. Call this graph G. Then we will see that the entropy of a simple random walk starting at any point on this graph grows at least linearly and yet the walk is recurrent. The example above was conveyed by Gady

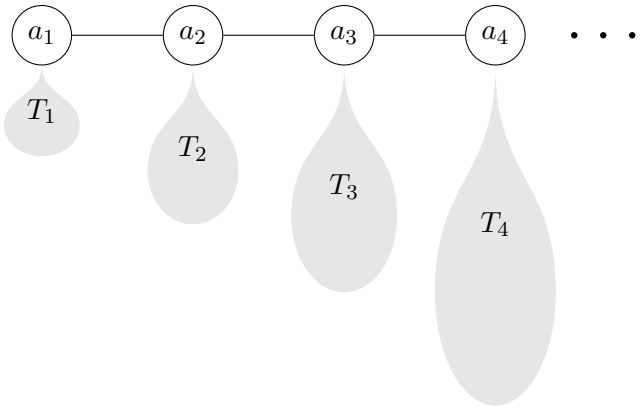

FIGURE 2.1. The Graph G

Kozma (personal communication, 2016) who suggested that the height of T_n be a tower of powers of 2, with n $2's$. We modified the height of the trees slightly to make the proofs easier.

THEOREM 2.6.1. *The simple random walk on G is recurrent.*

PROOF. Let $X_0, X_1, \ldots$ be a simple random walk on G that starts at a_1. For each $j \geq 1$, define A_j to be the set consisting of the vertex a_j and all the vertices of T_j, i.e.,

$$A_j = \{a_j\} \cup \{v \ : \ v \text{ is a vertex in } T_j\}.$$

We will analyze the successive times at which the random walk leaves A_j for some j. So, define $\tau(n)$ inductively by setting $\tau(0) = 0$ and

$$\tau(n) = \min\big\{t > \tau(n-1) \ : \ X_{\tau(n-1)} \in A_j \text{ and } X_t \notin A_j \text{ for some } j\big\}.$$

We will first show that $\tau(n) < \infty$ for all n, almost surely. Then, it is enough to show that $X_{\tau(0)}, X_{\tau(1)}, \ldots$ is recurrent. Note that $X_{\tau(n)}$ is supported on $\{a_1, a_2, \ldots\}$ for any n. So, to show that $\tau(n) < \infty$ for all n a.s., it is enough to show that $\tau(n+1) - \tau(n) < \infty$ a.s conditioned on $X_{\tau(n)} = a_j$, for every $j \geq 1$ and $n \geq 0$. Let $Y_0, Y_1, \ldots$ be a simple random walk on G starting at a_j for some $j \geq 1$. Let θ be the first time this walk exits A_j. then, by the strong Markov property,

$$\mathbb{P}\big(\tau(n+1) - \tau(n) < \infty \mid X_{\tau(n)} = a_j\big) = \mathbb{P}(\theta < \infty).$$

Since A_j is finite, $\mathbb{P}(\theta < \infty) = 1$. To see this, assume that the preceding probability is less than 1. This implies that there is a non-zero probability for the random walk to get trapped in A_j. This means that even for a simple random walk on the graph induced by $A_j \cup \{a_{j-1}\} \cup \{a_{j+1}\}$, that starts at a_j, there is a non-zero probability for the walk to get trapped in A_j. That is, this random walk doesn't hit a_{j+1} with probability 1. This contradicts the fact that random walks on finite connected graphs are always recurrent.

Next, observe that since $\tau(n)'s$ are stopping times, by the strong Markov property of the original chain, $X_{\tau(0)}, X_{\tau(1)}, \ldots$ is a Markov chain. We will show that in fact it is the simple random walk on the path $a_1a_2a_3\ldots$. Clearly, $\mathbb{P}(X_{\tau(n+1)} = a_2 \mid X_{\tau(n)} = a_1) = 1$, since a_1a_2 is the only edge between A_1 and A_2. For $j \geq 2$, just as above, let $Y_0, Y_1, \ldots$ be the simple random walk on G starting at a_j and let θ be the first time this random walk exits A_j. Then, by the strong Markov property,

$$\mathbb{P}\big(X_{\tau(n+1)} = b \mid X_{\tau(n)} = a_j\big) = \mathbb{P}(Y_\theta = b).$$

For $b \neq a_{j-1}, a_{j+1}$, we have $\mathbb{P}(Y_\theta = b) = 0$. Note that $Y_{\theta-1} = a_j$ always. If we condition on the event $\theta = m$ and $Y_{m-1} = a_j$ for any $m \geq 1$, then we get

$$\mathbb{P}\big(Y_\theta = a_{j+1} \mid \theta = m, Y_{m-1} = a_j\big) = \frac{\mathbb{P}(Y_{m-1} = a_j) \cdot 1/3}{\mathbb{P}(Y_{m-1} = a_j) \cdot (1/3 + 1/3)} = \frac{1}{2}.$$

This shows that $\mathbb{P}(Y_\theta = a_{j+1}) = 1/2$ and hence $X_{\tau(0)}, X_{\tau(1)}, \ldots$ is a simple random walk on the infinite path $a_1a_2\ldots$. It is well known that this walk is recurrent. This combined with the fact that $\tau(n)$ are all finite with probability 1, completes the proof. □

It remains to be shown that for any simply random walk on G, the entropy grows at least linearly. What we will show is that the random walk starting at a_1 has an entropy growth that is at least linear. From this it follows that for any starting position the entropy growth is at least linear. We will need the following lemma to prove these claims.

LEMMA 2.6.1. *Let T be an infinite binary tree with root r and let $Z_0, Z_1, \ldots$ be the simple random walk on T with $Z_0 = r$. Also, let E_n be the entropy of Z_n. Then for every n,*

$$E_n \geq \frac{\log 2}{3} n. \tag{2.26}$$

PROOF. Let $T = (V', E')$ and $L_n = \{x \in V' \ : \ d(r, x) = n\}$ where $d(\cdot, \cdot)$ is the graph distance. Let

$$p_n(i) = \mathbb{P}(Z_i \in L_n).$$

By symmetry it follows that for any $x \in L_n$,

$$\mathbb{P}(Z_i = x) = \frac{p_n(i)}{2^n}.$$

So from the above equation, we get

$$\begin{aligned} E_i &= \sum_{x \in V'} -\mathbb{P}(Z_i = x) \log \big(\mathbb{P}(Z_i = x)\big) = \sum_{n=0}^{\infty} \sum_{x \in L_n} -\mathbb{P}(Z_i = x) \log \big(\mathbb{P}(Z_i = x)\big) \\ &= \sum_{n=0}^{\infty} \sum_{x \in L_n} -\frac{p_n(i)}{2^n} \log \left(\frac{p_n(i)}{2^n}\right) = \sum_{n=0}^{\infty} -2^n \frac{p_n(i)}{2^n} \log \left(\frac{p_n(i)}{2^n}\right) \\ &= \sum_{n=0}^{\infty} n p_n(i) \log 2 - \sum_{n=0}^{\infty} p_n(i) \log p_n(i). \end{aligned}$$

Note that $0 \leq p_n(i) \leq 1$ and hence $-p_n(i) \log p_n(i) \geq 0$ and we have

$$E_i \geq \sum_{n=0}^{\infty} n p_n(i) \log 2 = \mathbb{E}\big(d(r, Z_i)\big) \log 2.$$

$\{d(r, Z_i)\}_{i \geq 1}$ is simply a biased random walk on $\mathbb{Z}$ starting at 0, with the additional condition that the random walker goes right with probability one when $d(r, Z_j) = 0$. That is, $\{d(r, Z_i)\}_{i \geq 1}$ is a biased random walk on $\mathbb{Z}$ that is reflected at 0. Therefore, the expectation of this random walk is least $\mathbb{E}\big[\sum_{n=1}^{i} Y_i\big]$, where $\mathbb{P}(Y_i = 1) = 2/3$ and $\mathbb{P}(Y_i = -1) = 1/3$. Hence we have

$$\mathbb{E}\big(d(r, Z_i)\big) \geq i(2/3 - 1/3) = i/3.$$

This proves the lemma. □

Recall that $B_n - 1$ is the height of T_n and r_n is its root. We will be needing the following lemma which will help us prove that the simple random walk on G starting at a_1 has linear entropy growth. What the lemma below tells us is that a random walk starting at a_1 spends most of its time in a large tree T_n, with a probability bounded away from 0.

LEMMA 2.6.2. *Let $N \geq 8$ and let $X_0, X_1, \ldots$ be the simple walk on G that starts at a_1. For any vertex b, let τ_b the first time the random walk hits b. Then, there exists a positive integer n such that $2N < B_n$ and*

$$\sum_{j \leq N/2} \mathbb{P}\big(\tau_{r_n} = j, X_t \in T_n \text{ for all } j \leq t \leq N\big) \geq 1/108.$$

PROOF. Define

$$n' = \min\left\{k \ : \ 2k \sum_{i=1}^{k} 2^{B_i} > N/8\right\}.$$

Set $n = n' + 1$. Then note that since B_i is non-decreasing in i,

$$B_n = B_{n'+1} = 32(n')^2 2^{B_{n'}} \geq 16 \cdot 2n' \sum_{i=1}^{n'} 2^{B_i} > 16(N/8) = 2N.$$

Define the finite graph H to be the subgraph of G that is induced by the vertices $a_1, a_2, \ldots, a_{n'}$ and all the vertices of $T_1, T_2, \ldots, T_{n'-1}$. Let's count the number of edges in this graph so that we can apply the commute time identity to the vertices a_1 and $a_{n'}$. T_j has height $B_j - 1$ and hence contains $2(2^{B_j - 1} - 1)$ edges. There are $n' - 1$ edges between $a_1, a_2, \ldots, a_{n'}$. For every $1 \leq j \leq n' - 1$ there is one edge between a_j and r_j. Therefore,

$$m_e = \text{Number of edges in } H = \sum_{i=1}^{n'-1} 2(2^{B_i - 1} - 1) + (n' - 1) + (n' - 1) = \sum_{i=1}^{n'-1} 2^{B_i}.$$

Note that by the minimality of n',

$$2(n' - 1) \sum_{i=1}^{n'-1} 2^{B_i} \leq N/8.$$

Let $\tau'_{a_{n'}}$ be the time a simple random walk starting at a_1 in the graph H takes to hit $a_{n'}$ for the first time. Note that since the random walker cannot leave H before hitting $a_{n'}$, we can conclude that $\tau_{n'} = \tau'_{n'}$ in distribution. Now, let $\tau'_{a_1 a_{n'}}$ be the commute time between a_1 and $a_{n'}$ in H, i.e.,

$$\tau'_{a_1 a_{n'}} = \min\{t \geq \tau'_{n'} \ : \ X'_t = a_1\},$$

where $(X'_t)_{t\geq 0}$ is the simple random walk on H that starts at a_1. Putting all this together and using the commute time identity, we get

$$\begin{aligned}
\mathbb{E}[\tau_{a_{n'}}] &= \mathbb{E}[\tau'_{a_{n'}}] \leq \mathbb{E}[\tau'_{a_1 a_{n'}}] \\
&= 2m_e \mathcal{R}(a_1 \leftrightarrow a_{n'} \text{ in } H) \qquad (2.27) \\
&= 2(n'-1)m_e \\
&= 2(n'-1)\sum_{i=1}^{n'-1} 2^{B_i} \leq N/8, \qquad (2.28)
\end{aligned}$$

where $\mathcal{R}(a_1 \leftrightarrow a_{n'} \text{ in } H)$ denotes the resistance between a_1 and $a_{n'}$ in H. The equality (2.27) is known as the commute time identity which was first proved in [**10**]. Let T be an infinite binary tree with root r. Let x be the vertex below r in T. Let $\widetilde{\tau}_r$ be the time it takes for a simple random walk starting at x to hit r for the first time. Then, it is well known that

$$\mathbb{P}(\widetilde{\tau}_r < \infty) = 1/2.$$

Let y be one of the two vertices connected to r_n in the tree T_n. Let $Z_0, Z_1, \ldots$ be the simple random walk on G such that $Z_0 = y$. Let $\bar{\tau}_{r_n}$ be the first time this walk hits the root r_n. Note that since $B_n > 2N$ and $N \geq 8$, it follows that $\text{height}(T_n) - 1 = B_n - 2 > N$. This means that the walk $Z_0, Z_1, \ldots$ cannot reach the leaves of T_n within time N. Hence, we have

$$\mathbb{P}(\bar{\tau}_{r_n} \leq N) = \mathbb{P}(\widetilde{\tau}_r \leq N) \leq \mathbb{P}(\widetilde{\tau}_r < \infty) = 1/2.$$

This means that

$$\mathbb{P}(\bar{\tau}_{r_n} > N) \geq 1 - 1/2 = 1/2.$$

So, we finally get the following

$$\mathbb{P}(Z_t \in T_n \text{ for } 0 \leq t \leq N) \geq \mathbb{P}(\bar{\tau}_{r_n} > N) \geq 1/2. \qquad (2.29)$$

We are now in a position to prove the lemma. Let's focus our attention back to the random walk on G starting at a_1. After hitting $a_{n'}$, the walk can transition to $a_{n'+1} = a_n$ with probability $1/3$. It can then transition to r_n with probability $1/3$ followed by a transition to y with probability $1/3$,

where y as defined above is one of two vertices in T_n that are connected to the root. After reaching y, by (2.29), the walk can stay in the tree T_n for the next N steps with probability at least $1/2$. Therefore, we get for $j \geq 2$

$$\begin{aligned}
&\mathbb{P}\big(\tau_{r_n} = j, X_t \in T_n \text{ for } j \leq t \leq N\big) \\
\geq &\mathbb{P}\big(\tau_{a_{n'}} = j-2, X_{j-1} = a_n, X_j = r_n, X_{j+1} = y, X_t \in T_n \text{ for } j+1 \leq t \leq j+N+1\big) \\
\geq &\mathbb{P}\big(\tau_{a_{n'}} = j-2\big)\frac{1}{3}\frac{1}{3}\frac{1}{3}\frac{1}{2} = \frac{1}{54}\mathbb{P}\big(\tau_{a_{n'}} = j-2\big). \qquad (2.30)
\end{aligned}$$

For $j = 1$ or $j = 0$, $\mathbb{P}\big(\tau_{r_n} = 0\big) = 0$. $N/4 + 2 \leq N/2$ since $N \geq 8$, hence we have

$$\begin{aligned}
\sum_{j \leq N/2} \mathbb{P}\big(\tau_{r_n} = j, X_t \in T_n \text{ for all } j \leq t \leq N\big) &\geq \sum_{j \leq N/4+2} \mathbb{P}\big(\tau_{r_n} = j, X_t \in T_n \text{ for all } j \leq t \leq N\big) \\
&\geq \sum_{j \leq N/4+2} \frac{1}{54}\mathbb{P}\big(\tau_{a_{n'}} = j-2\big) \quad \text{from (2.30)} \\
&= \frac{1}{54}\Big(1 - \mathbb{P}\big(\tau_{a_{n'}} > N/4\big)\Big) \\
&\geq \frac{1}{54}\left(1 - \frac{\mathbb{E}(\tau_{a_{n'}})}{N/4}\right) \quad \text{by Markov's inequality} \\
&\geq \frac{1}{54}\left(1 - \frac{N/8}{N/4}\right) \quad \text{from (2.28)} \\
&= \frac{1}{108}.
\end{aligned}$$

□

he above lemma combined with the following fact about entropy will enable us to prove that the entropy growth for a walk starting at a_1 is linear.

LEMMA 2.6.3. *Let Z be a random variable supported on a finite set $\Omega = \{x_1, \ldots, x_k\}$. Let $E_1, E_2, \ldots, E_m$ be disjoint events measureable with respect to Z. Let $\mathcal{P}$ be the distribution of Z and let $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_m$ be the distributions of Z conditioned on $E_1, E_2, \ldots, E_m$ respectively. For a distribution $\mathcal{P}'$ on Ω, use $\mathcal{H}(\mathcal{P}')$ to denote its entropy. Then,*

$$\mathcal{H}(\mathcal{P}) \geq \sum_{i=1}^{m} \mathbb{P}(E_i)\mathcal{H}(\mathcal{P}_i).$$

PROOF. This lemma is a direct consequence of the fact that the function $\mathscr{D} : [0,1] \to \mathbb{R}$ defined by $\mathscr{D}(x) = -x\ln(x)$ for $x \neq 0$ and $\mathscr{D}(0) = 0$, is concave. Let $E_{m+1} = (E_1 \cup E_2 \cup \cdots \cup E_m)^c$ and let $\mathscr{P}_{m+1} \in \mathbb{R}^{|\Omega|}$ be the distribution of Z conditioned on E_{m+1}. First, note that for any probability distribution $\mathscr{P}'$ on $\Omega = \{x_1, \ldots, x_k\}$, by definition of the entropy,

$$\mathscr{H}(\mathscr{P}') = \sum_{i=1}^{k} \mathscr{D}(\mathscr{P}'(x_i)).$$

Next, note that by the concavity of $\mathscr{D}(\cdot)$, for any $i \in \{1, \ldots, k\}$

$$\mathscr{D}\Big(\sum_{j=1}^{m+1} \mathbb{P}(E_j)\mathscr{P}_j(x_i)\Big) \geq \sum_{j=1}^{m+1} \mathbb{P}(E_j)\mathscr{D}\big(\mathscr{P}_j(x_i)\big). \tag{2.31}$$

So, finally we have

$$\begin{aligned}
\mathscr{H}(\mathscr{P}) = \sum_{i=1}^{k} \mathscr{D}(\mathscr{P}(x_i)) &= \sum_{i=1}^{k} \mathscr{D}\Big(\sum_{j=1}^{m+1} \mathbb{P}(E_j)\mathscr{P}_j(x_i)\Big) \\
&\geq \sum_{i=1}^{k}\sum_{j=1}^{m+1} \mathbb{P}(E_j)\mathscr{D}\big(\mathscr{P}_j(x_i)\big) \quad \text{from (2.31)} \\
&= \sum_{j=1}^{m+1} \mathbb{P}(E_j) \sum_{i=1}^{k} \mathscr{D}\big(\mathscr{P}_j(x_i)\big) \\
&= \sum_{j=1}^{m+1} \mathbb{P}(E_i)\mathscr{H}\big(\mathscr{P}_i\big).
\end{aligned}$$

Since the above sum consists of non-negative terms, our proof is complete. □

THEOREM 2.6.2. *Let $X_0, X_1, \ldots$ be the simple random walk on G with $X_0 = a_1$. Let E_n be the entropy of X_n. Then there exists a constant $\gamma > 0$ such that for every $n \geq 0$,*

$$E_n \geq \gamma n.$$

PROOF. Fix $N \geq 8$ and let n be as in Lemma 2.6.2. Let $\mathcal{R}$ be the support of X_N. Let $\mathscr{P} \in \mathbb{R}^{|\mathcal{R}|}$ be the distribution of X_N viewed as a vector. For any probability distribution $\mathscr{P}' \in \mathbb{R}^{|\mathcal{R}|}$, let $\mathscr{H}(\mathscr{P}')$ denote its entropy. Our aim is to apply Lemma 2.6.3 above. Let

$$E_j = \{\tau_{r_n} = j, X_t \in T_n \text{ for all } j \leq t \leq N\}.$$

Let $\mathscr{P}_j$ be the distribution of X_N conditioned on E_j. Note that if we condition on the event that a random walk starting at r_n stays in T_n for m steps where $m \leq \text{height}(T_n)$ then the conditional distribution of this walk for the first m steps is the same as the distribution for the first m steps of a simple random walk that starts at the root of an infinite binary tree. Therefore the distribution of X_N conditioned on E_j is no different from the distribution of the $(N-j)^{th}$ step of a simple random walk starting at the root of an infinite binary tree. Hence, from Lemma 2.6.1 we get that for $j \leq N/2$,

$$\mathscr{H}(\mathscr{P}_j) \geq \frac{\log 2}{3}(N-j) \geq \frac{\log 2}{3}\frac{N}{2}. \tag{2.32}$$

The events E_j are disjoint since the events $\{\tau_{r_n} = j\}$ are disjoint. Hence we can now apply Lemma 2.6.3 and Lemma 2.6.2 to conclude that

$$\begin{aligned} E_N = \mathscr{H}(\mathscr{P}) &\geq \sum_{j \leq N/2} \mathbb{P}(E_j)\mathscr{H}(\mathscr{P}_j) \\ &\geq \frac{\log 2}{3}\frac{N}{2} \sum_{j \leq N/2} \mathbb{P}(E_j) \quad \text{from (2.32)} \\ &\geq \frac{\log 2}{3}\frac{N}{2}\frac{1}{108} \quad \text{from Lemma 2.6.2.} \end{aligned}$$

We have shown that for $N \geq 8$,

$$E_N \geq \frac{\log 2}{648} N.$$

$E_0 = 0$, hence by choosing a small enough $\gamma > 0$ we can ensure that for all $N \geq 0$

$$E_N \geq \gamma N.$$

□

Corollary 2.6.1. *Let $X_0, X_1, \ldots$ be a simple random walk on G such that $X_0 = x$ and let E_n be the entropy of X_n. Then there exists a constant $\gamma' > 0$ such that*

$$E_n \geq \gamma' n.$$

PROOF. Let $m = d(a_1, x)$ and let E be the event

$$E = \{X_m = a_1\}.$$

Then $\mathbb{P}(E) > 0$. Now, for $N \geq 2m$ let $\mathscr{P}$ be the distribution of X_N and let $\mathscr{P}'$ be the distribution of X_N conditioned on E. Then by Theorem 2.6.2 and Lemma 2.6.3 we have,

$$\mathscr{H}(\mathscr{P}) \geq \mathbb{P}(E)\mathscr{H}(\mathscr{P}') \geq \mathbb{P}(E)\frac{\log 2}{648}(N - m) \geq \mathbb{P}(E)\frac{\log 2}{648}\frac{N}{2}.$$

This holds for $N \geq m$ and moreover $E_0 = 0$. Therefore, there exists a $\gamma' > 0$ possibly depending on m, such that

$$E_n \geq \gamma' n$$

for every $n \geq 0$. □

Bibliography

[1] J. Alwen, Y. Dodis, M. Naor, G. Segev, S. Walfish, and D. Wichs, *Public-key encryption in the bounded-retrieval model*, in Advances in Cryptology - EUROCRYPT 2010, 29th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Monaco / French Riviera, May 30 - June 3, 2010. Proceedings, H. Gilbert, ed., vol. 6110 of Lecture Notes in Computer Science, Springer, 2010, pp. 113–134.

[2] J. Alwen, Y. Dodis, and D. Wichs, *Leakage-resilient public-key cryptography in the bounded-retrieval model*, in Advances in Cryptology - CRYPTO 2009, 29th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 16-20, 2009. Proceedings, S. Halevi, ed., vol. 5677 of Lecture Notes in Computer Science, Springer, 2009, pp. 36–54.

[3] A. Avez, *Entropie des groupes de type fini*, C.R. Acad. Sci. Paris Sér. A-B, 275 (1972), pp. A1363–A1366.

[4] A. Avez, *Théorème de choquet-deny pour les groupes à croissance non exponentionelle*, C.R. Acad. Sci. Paris Sér. A, 275 (1974), pp. 25–28.

[5] A. Avez, *Croissance des groupes de type fini et fonctions harmoniques*, in Théorie Ergodique, Lecture Notes in Mathematics, Springer, Berlin, 1976, pp. 35–49.

[6] M. Bellare, A. Desai, E. Jokipii, and P. Rogaway, *A concrete security treatment of symmetric encryption*, in 38th Annual Symposium on Foundations of Computer Science, FOCS '97, Miami Beach, Florida, USA, October 19-22, 1997, IEEE Computer Society, 1997, pp. 394–403.

[7] M. Bellare, D. Kane, and P. Rogaway, *Big-key symmetric encryption: Resisting key exfiltration*, in Advances in Cryptology - CRYPTO 2016 - 36th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2016, Proceedings, Part I, M. Robshaw and J. Katz, eds., vol. 9814 of Lecture Notes in Computer Science, Springer, 2016, pp. 373–402.

[8] M. Bellare and P. Rogaway, *Random oracles are practical: A paradigm for designing efficient protocols*, in CCS '93, Proceedings of the 1st ACM Conference on Computer and Communications Security, Fairfax, Virginia, USA, November 3-5, 1993, D. E. Denning, R. Pyle, R. Ganesan, R. S. Sandhu, and V. Ashby, eds., ACM, 1993, pp. 62–73.

[9] D. Cash, Y. Z. Ding, Y. Dodis, W. Lee, R. J. Lipton, and S. Walfish, *Intrusion-resilient key exchange in the bounded retrieval model*, in Theory of Cryptography, 4th Theory of Cryptography Conference, TCC 2007, Amsterdam, The Netherlands, February 21-24, 2007, Proceedings, S. P. Vadhan, ed., vol. 4392 of Lecture Notes in Computer Science, Springer, 2007, pp. 479–498.

[10] A. K. Chandra, P. Raghavan, W. L. Ruzzo, R. Smolensky, and P. Tiwari, *The electrical resistance of a graph captures its commute and cover times*, Comput. Complexity, 6(4) (1989), pp. 312–340.

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, 2006.

[12] D. D. Crescenzo, R. J. Lipton, and S. Walfish, *Perfectly secure password protocols in the bounded retrieval model*, in Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings, S. Halevi and T. Rabin, eds., vol. 3876 of Lecture Notes in Computer Science, Springer, 2006, pp. 225–244.

[13] Y. Derriennic, *Quelques applications du théorème ergodique sous-additif*, Astérisque, 74 (1980), pp. 183–201.

[14] P. Diaconis and J. A. Fill, *Strong stationary times via a new form of duality*, The Annals of Probability, 18(4) (1990), pp. 1483–1522.

[15] S. Dziembowski, *Intrusion-resilience via the bounded-storage model*, in Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings, S. Halevi and T. Rabin, eds., vol. 3876 of Lecture Notes in Computer Science, Springer, 2006, pp. 207–224.

[16] S. Gouëzel, F. Mathéus, and F. Maucourant, *Sharp lower bounds for the asymptotic entropy of symmetric random walks*, Groups Geom. Dyn., 9(3) (2015), pp. 711–735.

[17] V. A. Kaimanovich and W. Woess, *Boundary and entropy of space homogeneous markov chains*, Ann. Probab., 30(1) (2002), pp. 323–363.

[18] F. Ledrappier, *Sharp estimates for the entropy*, in Proceedings of the International Meeting held in Frascati, July 1–10, 1991, M. A. Picardello, ed., 1992, pp. 281–288.

[19] D. Levin, Y. Peres, and E. Wilmer, *Markov Chains and Mixing Times*, American Mathematical Society, 2017.

[20] R. Lyons and Y. Peres, *Probability on Trees and Networks*, Cambridge University Press, 2016.

[21] B. Morris, *The mixing time of the thorp shuffle*, SIAM J. on Computing, 38(2) (2008), pp. 484–504.

[22] B. Morris, *Improved mixing time bounds for the thorp shuffle and l-reversal chain*, Annals of Probability, 37(2) (2009), pp. 453–477.

[23] B. Morris, *Improved mixing time bounds for the thorp shuffle*, Combinatorics, Probability and Computing, 22(1) (2013), pp. 118–132.

[24] B. Morris and Y. Peres, *Evolving sets, mixing and heat kernel bounds*, Probability Theory and Related Fields, 133(2) (2005), pp. 245–266.

[25] B. Morris, P. Rogaway, and T. Stegers, *How to encipher messages on a small domain*, in Advances in Cryptology - CRYPTO 2009, 29th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 16-20, 2009. Proceedings, S. Halevi, ed., vol. 5677 of Lecture Notes in Computer Science, Springer, 2009, pp. 286–302.

[26] USER125368, *An upper bound of binary entropy.* Mathematics Stack Exchange. URL: https://math.stackexchange.com/q/1432043 (version: 2015-09-14).

[27] A. M. VERSHIK AND V. A. KAIMANOVICH, *Random walks on groups: Boundary, entropy, uniform distribution*, Dokl. Akad. Nauk SSSR, 249(1) (1979), pp. 15–18.