Lossy Chains and Fractional Secret Sharing *

Yuval Ishai[†], Eyal Kushilevitz[‡], and Omer Strulovich[§]

Technion, Haifa, Israel
{yuvali,eyalk,omers}@cs.technion.ac.il

February 24, 2013

Abstract

Motivated by the goal of controlling the amount of work required to access a shared resource or to solve a cryptographic puzzle, we introduce and study the related notions of *lossy chains* and *fractional secret sharing*.

Fractional secret sharing generalizes traditional secret sharing by allowing a fine-grained control over the amount of uncertainty about the secret. More concretely, a fractional secret sharing scheme realizes a fractional access structure $f : 2^{[n]} \to [m]$ by guaranteeing that from the point of view of each set $T \subseteq [n]$ of parties, the secret is *uniformly* distributed over a set of $f(T)$ potential secrets. We show that every (monotone) fractional access structure can be realized. For *symmetric* structures, in which $f(T)$ depends only on the size of T, we give an efficient construction with share size $\mathrm{poly}(n, \log m)$.

Our construction of fractional secret sharing schemes is based on the new notion of *lossy chains* which may be of independent interest. A lossy chain is a Markov chain $(X_0, \ldots, X_n)$ which starts with a random secret X_0 and gradually loses information about it at a rate which is specified by a *loss function* g. Concretely, in every step t, the distribution of X_0 conditioned on the value of X_t should always be uniformly distributed over a set of size $g(t)$. We show how to construct such lossy chains efficiently for any possible loss function g, and prove that our construction achieves an optimal asymptotic information rate.

1 Introduction

In this work, we introduce and study two related notions: *lossy chains* and *fractional secret sharing*. We start by describing the latter.

Fractional secret sharing. Suppose that we wish to share a secret password between several parties, such that the largest subset of cooperating parties will be the first to guess the correct password. (Think of the password as a key which locks a physical or digital vault,

*This is the full version of [6].

[†]Supported by the European Research Council as part of the ERC project CaC (grant 259426), ISF grant 1361/10, and BSF grant 2008411.

[‡]Supported by ISF grant 1361/10 and BSF grant 2008411.

[§]Supported by ERC grant 259426.

where the number of guessing attempts measures the amount of work required for unlocking the vault.)

A simple solution that comes to mind is the following. If the password is a binary string of length k and we have $n \leq k$ parties, we can give each party one or more bits of the password. In this solution, a larger cooperating subset of parties will need a smaller expected number of attempts to guess the password than a smaller one. This solution achieves our goal, but is limited to specific parameters. For example, we cannot easily extend this method to $n > k$ parties, nor can we have finer control over the relative amount of expected work required by different subsets of parties.

Our goal is to find a solution which gives maximal control over the amount of information about the password revealed to each subset of parties. This motivates the notion of *fractional secret sharing*. In traditional secret sharing [10, 3, 7], each subset of n parties either has full information about the secret or has no information about the secret. Fractional secret sharing generalizes this notion by allowing a fine-grained control over the amount of uncertainty of each subset about a uniformly random secret. The uncertainty is specified by a *fractional access structure* $f : 2^{[n]} \to [m]$. A fractional secret sharing scheme realizing f should have the property that from the point of view of each set $T \subseteq [n]$ of parties, the secret is always *uniformly* distributed over a set of $f(T)$ potential secrets. Since adding a party to a subset cannot reduce the amount of available information, we assume f to be *monotone* in the sense that if $T \subseteq T'$ then $f(T') \leq f(T)$. This raises the following questions:

Can every (monotone) fractional access structure be realized? If so, how efficiently?

How to gradually forget. Motivated in part by the problem of fractional secret sharing, we introduce the related notion of *lossy chains*. A lossy chain is a Markov chain $(X_0, \ldots, X_n)$ which starts with a random secret X_0 and gradually loses information about it at a rate which is specified by a predefined loss function $g : [n] \to [m]$. More concretely, we require that for any $1 \leq i \leq n$ and any possible value x_i in the support of X_i, the distribution of the secret X_0, conditioned on the event that $X_i = x_i$, is uniform[1] over a set of size $g(i)$. (The identity of this set may depend on x_i.) In a similar manner to fractional access structures, we require that the loss function g be monotone, in the sense that for $i < j$ we have $g(i) < g(j)$. This raises the following questions:

Can every (monotone) loss function be realized? If so, how efficiently?

The Markov property of the chain (namely, the requirement that X_{i+1} be independent of $X_0, \ldots, X_{i-1}$ given X_i) is important for our motivating applications, as it rules out the possibility of combining several values X_i in order to learn more information than that implied by the "best" value X_i. Jumping ahead, this property will turn out to be crucial for the construction of fractional secret sharing from lossy chains.

Why uniform? An important aspect of our notions of fractional secret sharing and lossy chains is that they require each conditional distribution to always be *uniform* over a subset of potential secrets having a specified size. One could instead consider alternative definitions

[1]The uniformity requirement rules out simple solutions that are based on gradually adding independent random noise to the initial secret (cf. [4]), see further discussion below.

which only specify some measure of *entropy*, such as conditional Shannon entropy [11] or min-entropy [9], without further restricting the distribution. Insisting on a uniform distribution has several important advantages. First, a crude measure of uncertainty such as entropy is not informative enough to capture all relevant properties of a distribution. For instance, min-entropy determines the best probability of guessing the secret in the first attempt, but says little about the expected number of attempts until the secret is correctly guessed. Second, using the uniform distribution does not only give control over the expected number of attempts in an optimal guessing strategy, but it also minimizes the *variance* of the number of such attempts under the expectation constraint. (See Appendix A for a proof that the uniform distribution beats any other distribution in this respect.) Finally, in some scenarios it may be desirable to spread the point in time in which the secret is correctly guessed as evenly as possible (think of a password controlling a shared resource). This too is achieved optimally by the uniform distribution. We note that one could relax the requirement of uniformity to being statistically close to uniform. This is addressed in Appendix B.

1.1 Our Results

We obtain several positive and negative results about lossy chains and fractional secret sharing.

- We show that any monotone loss function $g : [n] \to [m]$ can be efficiently realized by a lossy chain $(X_0, \ldots, X_n)$ in which the bit-length of each X_i is at most $n \cdot \lceil \log m \rceil$. Moreover, we show this bound to be asymptotically tight by demonstrating the existence of a family of loss functions $g_{n,m} : [n] \to [m]$ for which some X_i must be $\Omega(n \log m)$ bits long. This asymptotic lower bound still holds even if we allow the conditional distributions to have negligible statistical distance from uniform. Settling for a constant statistical distance, the bit-length of each X_i can be $O(\log^2 m)$, independently of n.

- We show a general reduction of fractional secret sharing to lossy chains, which implies that every monotone fractional access structure $f : 2^{[n]} \to [m]$ can be realized. For the important case of *symmetric* structures, in which $f(T)$ depends only on the size of T, we get an efficient construction in which the share size of each party is at most $n \cdot \lceil \log \max\{n, m\} \rceil$.

1.2 Overview of Techniques

Recall that a lossy chain is a Markov chain $(X_0, \ldots, X_n)$, where X_0 is a random secret, and each step loses additional information about the secret. This loss is specified by a loss function $g : [n] \to [m]$, such that for each $1 \leq i \leq n$ and x_i in the support of X_i, the distribution of X_0 conditioned on $X_i = x_i$ is uniform over a set of size $g(i)$. (See Section 3.1 for a formal definition.)

As a simple warmup example, let X_0 be uniform over $\{0,1\}^n$, and let X_i include the first $n-i$ bits of X_0. In this case, X_0 conditioned on $X_i = x_i$ is distributed uniformly over a set of size 2^i. Thus, this lossy chain realizes the loss function $g(i) = 2^i$. This simple approach only works for a loss function g which is increasing exponentially, and can be generalized only to loss functions g such that $g(i)$ divides $g(i+1)$.

The following alternative approach works for any monotone loss function $g : [n] \to [m]$, where without loss of generality $g(n) = m$:

1. Pick x_0 uniformly from $[m]$.
2. For $i = 1, \ldots, n$, pick (a set) x_i uniformly at random from all subsets of $[m]$ of size $g(i)$ containing x_{i-1}.
3. Output $(x_0, x_1, \ldots, x_n)$.

Intuitively, this method starts from a set $\{x_0\}$ containing only the correct secret, and in each step adds $g(i) - g(i-1)$ new random "distractors" from $[m]$. This allows us to realize a lossy chain for any loss function. However, storing or sending the values of such a chain may be infeasible when m is large (e.g., think of m as the number of possible passwords). It is therefore desirable to get a solution in which the bit-length of each X_i grows logarithmically with m instead of linearly with m.

A natural approach is to limit the subsets represented by X_i to only be discrete intervals of the form $[j, k] = \{j, j+1, \ldots, k\}$, where $1 \leq j \leq k \leq m$. Unfortunately, this simple modification of the previous construction fails to satisfy the uniform conditional distribution property. More concretely, given an interval $[j, k]$ for X_i, the probability of the secret X_0 being in the middle of the interval will be higher than in the edges of the interval. To avoid this problem, we employ *cyclic* intervals. Intuitively, given an arbitrary ordered set, a cyclic interval can "cycle" through the end back to the start of the set. Using recursive nesting of such cyclic intervals, we construct a lossy chain for any loss function while keeping the support of each X_i small. We describe our results for lossy-chains in Section 3. We present the above construction in Section 3.2, and we establish the optimality of this construction in Section 3.3 by using some basic linear algebraic properties of the probability vectors associated with a lossy chain. Positive and negative results for the statistical relaxation of lossy chains are given in Appendix B.

Finally, in Section 4 we describe the reduction of fractional secret sharing to lossy chains. Recall that a fractional secret sharing scheme realizes a fractional access structure $f : 2^{[n]} \to [m]$ by ensuring that from the point of view of each set $T \subseteq [n]$ of parties, the secret is *uniformly* distributed over a set of $f(T)$ potential secrets. (See Section 4.1 for a formal definition.) In the case of a symmetric structure f, where $f(T)$ depends only on the size of T, we can use the following natural construction: let $g(i) = f([n-i])$ and let $(X_0, \ldots, X_n)$ be a lossy chain realizing g. A fractional secret sharing scheme realizing f can be obtained by using a threshold secret sharing scheme (such as Shamir's scheme [10]) to distribute the value of each X_i between the n parties with reconstruction threshold $n - i$. Any set T of t parties will be able to reconstruct the values $X_{n-t}, \ldots, X_n$, which by the Markov property contain the same information about the secret X_0 as X_{n-t}. By the definition of g, the distribution of X_0 conditioned on the value of X_{n-t} is uniform over a set of size $f(T)+1$, as required. The above construction can be generalized to arbitrary fractional access structures. However, similarly to traditional secret sharing, the complexity of the general construction may be exponential in the number of parties.

Related work. The notion of fractional secret sharing can be viewed as a restricted instance of *non-perfect* secret sharing (also referred to as *ramp secret sharing*). While in standard (perfect) secret sharing schemes each set of players should either be able to fully reconstruct the secret or alternatively should learn nothing about it, in non-perfect secret sharing there is also a third kind of sets that may learn partial information about the secret. Non-perfect

schemes were proposed mainly for the reason of improving the efficiency of secret sharing by reducing the size of the shares. Unlike fractional secret sharing, in non-perfect secret sharing there is no requirement on the type of partial information available to the third kind of subsets. For works on non-perfect secret sharing, see [2, 12, 8, 5] and references therein.

2 Preliminaries

Notation. We let $[n]$ denote the set of integers $\{1, 2, \ldots, n\}$. We use $\log n$ to denote $\log_2 n$. For a random variable X, we let $\mathsf{supp}(X)$ denote the support set of X, that is, the set of values which X may take with nonzero probability. The support set of a real-valued vector is the set of coordinates in which it takes nonzero values.

Markov chains. A Markov chain is a sequence of random variables such that the distribution of each variable in the sequence depends only on the value of the previous variable. Formally:

Definition 1. (Markov chain) Let $\bar{X} = (X_0, X_1, \ldots, X_n)$ be a sequence of jointly distributed random variables. We say that $\bar{X}$ is a *Markov chain* if for every $i \in [n]$ and for any sequence of values $x_0 \in \mathsf{supp}(X_0), \ldots, x_i \in \mathsf{supp}(X_i)$,

$$\Pr\left[X_i = x_i | X_{i-1} = x_{i-1}\right] = \Pr\left[X_i = x_i | X_{i-1} = x_{i-1}, \ldots, X_0 = x_0\right].$$

In general, Markov chains can be defined as infinite sequences of random variables with infinite support size. However, in this work we will only consider finite Markov chains.

The above definition is equivalent to requiring that for any i and x_i in the support set of X_i, the random variables $(X_1, \ldots, X_{i-1})$ and $(X_{i+1}, \ldots, X_n)$ are independent conditioned on $X_i = x_i$. The symmetry of the above conditional independence requirement implies the following "reversibility" property of Markov chains (see [1, p. 215] for a formal proof).

Fact 1. *If $\bar{X} = (X_0, X_1, \ldots, X_n)$ is a Markov chain, then so is $\bar{X}^R = (X_n, X_{n-1}, \ldots, X_0)$.*

3 Lossy Chains

In this section we define our new notion of a lossy chain (Section 3.1), present an efficient construction of lossy chains (Section 3.2), and prove a lower bound on their efficiency (Section 3.3).

3.1 Definitions and Basic Properties

A lossy chain is a Markov chain in which the information loss about the initial value is fully specified by a *loss function*. We start by defining the latter.

Definition 2. (Loss function) A *loss function* is a monotone increasing function $g : [n] \to [m]$. That is, for every $1 \leq i < j \leq n$ we have $g(i) < g(j)$.

We now turn to define lossy chains.

Definition 3. (Lossy chain) Let $g : [n] \to [m]$ be a loss function, and let $\bar{X} = (X_0, X_1, \ldots, X_n)$ be a sequence of random variables. We say that $\bar{X}$ is a *lossy chain realizing g* if the following conditions hold:

- $\bar{X}$ is a Markov chain, and
- for every $i \in [n]$ and every x_i in the support of X_i, the distribution of X_0 conditioned on $X_i = x_i$ is uniform over a set of size $g(i)$.

Our goal is to construct lossy chains in which each value can be succinctly described. To this end we use the following measure of efficiency.

Definition 4. (Information rate) Let $\bar{X} = (X_0, X_1, \ldots, X_n)$ be a lossy chain. The *information rate* of $\bar{X}$ is defined as

$$\rho(\bar{X}) = \min_{0 \le i \le n} \frac{\log g(n)}{\log |\mathsf{supp}(X_i)|}.$$

It will be convenient to assume that in a lossy chain realizing $g : [n] \to [m]$, the initial value X_0 is uniformly distributed over a set of size $g(n)$, and X_n has support set of size 1. The following claim shows that this assumption is without loss of generality: any lossy chain realizing g can be converted into a canonical form that has this property and has the same or better information rate.

Claim 1. (Canonical lossy chain) *Let $g : [n] \to [m]$ be a loss function and let $\bar{X} = (X_0, X_1, \ldots, X_n)$ be a lossy chain realizing g. Let x_n be an arbitrary element in the support of X_n. Let $\bar{X}' = (X'_0, X'_1, \ldots, X'_n)$ be the joint distribution defined by*

$$\Pr[\bar{X}' = (x'_0, x'_1, \ldots, x'_n)] = \Pr[\bar{X} = (x'_0, x'_1, \ldots, x'_n) \mid X_n = x_n].$$

Then $\bar{X}'$ is a lossy chain realizing g. Moreover, X'_0 is uniform over a set of size $g(n)$ and $\mathsf{supp}(X'_i) \subseteq \mathsf{supp}(X_i)$ for $0 \le i \le n$.

Proof. To see that $\bar{X}'$ is a Markov chain, note that if $(X_1, \ldots, X_{i-1})$ and $(X_{i+1}, \ldots, X_n)$ are independent when conditioned on $X_i = x_i$, then $(X_1, \ldots, X_{i-1})$ and $(X_{i+1}, \ldots, X_{n-1})$ are independent when conditioned on $X_i = x_i$ and $X_n = x_n$.

We now show that $\bar{X}'$ realizes g. For this, it suffices to show that for any $i \in [n]$ and any $x'_i \in \mathsf{supp}(X'_i)$, the distribution of X'_0 conditioned on $X'_i = x'_i$ is identical to the distribution of X_0 conditioned on $X_i = x'_i$. Indeed, for any x'_0 we have

$$\begin{aligned} \Pr\left[X'_0 = x'_0 | X'_i = x'_i\right] &= \Pr\left[X_0 = x'_0 | X_i = x'_i, X_n = x_n\right] \\ &= \Pr\left[X_0 = x'_0 | X_i = x'_i\right], \end{aligned}$$

where the first equality follows from the definition of $\bar{X}'$ and the second from the conditional independence property of Markov chains. Finally, the fact that X'_0 is uniform over a set of size $g(n)$ follows immediately from Definition 3 and the fact that $\mathsf{supp}(X'_i) \subseteq \mathsf{supp}(X_i)$ follows from X'_i being a restriction of X_i to a conditional space. □

3.2 An Efficient Construction

In the Introduction, we have seen a simple general construction of a lossy chain realizing $g : [n] \to [m]$ whose information rate is $\tilde{\Theta}(1/m)$. This construction may be infeasible for large values of m. In this section, we show how the rate can be improved to $1/n$.

We first recall the scheme described in the Introduction. Given $g : [n] \to [m]$ where (without loss of generality) $g(n) = m$, the lossy chain is computed as follows.

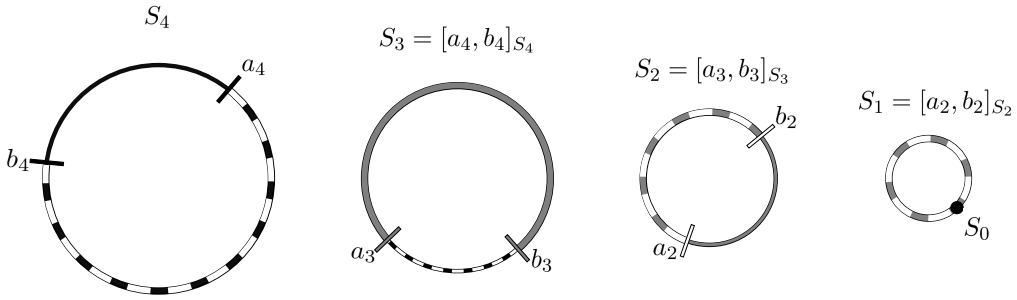

Figure 1: The cyclic interval from a_4 to b_4 is taken from a set S_4 with a cyclic order to create S_3. Then S_2 is created as a subset of S_3 by taking another cyclic interval. This goes on until S_0, the starting value, is chosen from S_1.

1. Pick x_0 uniformly from $[m]$.
2. For $i = 1, \ldots, n$, pick a set x_i uniformly at random from all subsets of $[m]$ of size $g(i)$ containing x_{i-1}.
3. Output $(x_0, x_1, \ldots, x_n)$.

This chain is inefficient in that it requires to store arbitrary subsets of $[m]$. In order to obtain a more efficient variant of this construction, we restrict these subsets to be nested *cyclic intervals*.

Definition 5. (Cyclic interval) Let $S = \{e_0, \ldots, e_{m-1}\}$ be a linearly ordered set, where $e_0 < e_1 < \ldots < e_{m-1}$. For any two integers $a, b \in \{0, \ldots, m-1\}$, the *cyclic interval from a to b over S,* denoted $[a, b]_S$, is defined by:

$$[a, b]_S = \begin{cases} \{e_a, \ldots, e_b\} & a \leq b \\ \{e_a, \ldots, e_{m-1}\} \cup \{e_0, \ldots e_b\} & a > b \end{cases}.$$

Note that for a given size k, there are exactly $|S|$ distinct nested intervals of size k in S, one for each starting point a. The algorithm for generating the lossy chain iteratively generates subsets S_i of size $g(i)$ for every $i \in [n]$ in *decreasing order*, where each subset S_i is a random cyclic interval in S_{i+1}. See Figure 1 for a visual illustration. A precise description of the algorithm is given in Figure 2.

We now prove that the output of the "Cyclic Intervals" algorithm from Figure 2 forms a lossy chain realizing g. In the following, we denote by $\bar{X} = (X_0, \ldots, X_n)$ the joint distribution of the output. We start by showing that the output indeed forms a Markov chain.

Lemma 1. *The output distribution $(X_0, \ldots, X_n)$ forms a Markov chain.*

Proof. For $1 \leq i \leq n$, the output X_{i-1} is sampled based on X_i alone. This implies that $(X_n, \ldots, X_0)$ is a Markov chain and, by Fact 1, we have that $(X_0, \ldots, X_n)$ is also a Markov chain. □

"Cyclic Intervals" lossy chain

Input: A loss function $g : [n] \to [m]$.

Algorithm:

1. $S_n \leftarrow [g(n)]$
2. For $i = n-1, ..., 1$
   (a) Pick $a_i \in \{0, ..., g(i) - 1\}$ uniformly at random
   (b) $b_i \leftarrow (a_i + g(i) - 1) \bmod g(i+1)$
   (c) set $S_i = [a_i, b_i]_{S_{i+1}}$
3. Pick x_0 uniformly at random from S_1

Output: $\bar{x} = (x_0, S_1, S_2, \ldots, S_n)$

Figure 2: Lossy chain obtained via nested cyclic intervals

Lemma 2. *The chain $\bar{X}$ realizes the loss function g.*

Proof. We prove that for any $1 \leq i \leq n$ and any $S_i \in \mathsf{supp}(X_i)$, the distribution of X_0 conditioned on the event $X_i = S_i$ is distributed uniformly over S_i. Since $|S_i| = g(i)$ the lemma will follow.

The above claim intuitively follows by symmetry. We formally prove it by induction on i. The case $i = 1$ follows directly from the algorithm's description. Suppose the claim holds for i, and let S_{i+1} be in the support of X_{i+1}. We need to prove that X_0 conditioned on $X_{i+1} = S_{i+1}$ is uniformly distributed over S_{i+1}. Indeed, when $X_{i+1} = S_{i+1}$ the choice of the output x_0 can be viewed as resulting from the following two step process:

1. Pick S_i as a random cyclic interval in S_{i+1} of size $g(i)$.
2. Pick x_0 from the distribution of X_0 conditioned on $X_i = S_i$.

The choice of S_i in the first step guarantees that each $x \in S_{i+1}$ has an equal probability to be in S_i. By the induction's hypothesis, the second step picks x_0 uniformly at random from S_i. Combining the two steps, x_0 is uniformly distributed over S_{i+1}, as required. □

Using the above two lemmas, we obtain the main theorem of this section.

Theorem 1. *For any loss function $g : [n] \to [m]$, there is a lossy chain realizing g whose information rate is at least $\frac{1}{n-1}$.*

Proof. Lemma 1 and Lemma 2 imply that the algorithm from Figure 2 is a lossy chain realizing g. The bound on the information rate follows from the fact that each S_i can be fully specified using the sequence $(a_{n-1}, \ldots, a_1)$, where $0 \leq a_i < g(i) \leq g(n)$ for all i, and from the fact that $|\mathsf{supp}(X_0)| = g(n)$. □

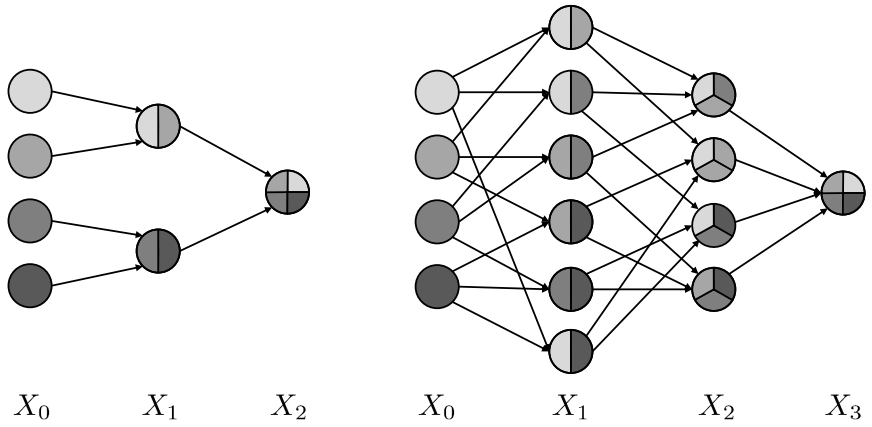

Figure 3: On the left, a states graph for a simple lossy chain realizing $g(i) = 2^i$ with 4 possible starting values. On the right, a lossy chain realizing $g(i) = i + 1$, also with 4 possible starting values.

Remark 1. (On computational efficiency) The description in Figure 2 does not address the question of how the sets S_i are represented and how one can efficiently enumerate the elements of S_i or sample from S_i. To this end, we note that if we modify the algorithm such that X_i contains the representation of S_i by the sequence $(a_{n-1}, \ldots, a_i)$, the resulting chain still realizes g (namely, the additional information provided by this sequence does not change the distribution of X_0 conditioned on S_i). Moreover, the information rate of this (slightly redundant) representation of the sets S_i is still lower bounded by $1/(n-1)$. *See Appendix C for efficient algorithms supporting this representation.*

3.3 A Negative Result

In this section, we establish a limitation on the information rate of lossy chains, showing that the cyclic intervals construction cannot be asymptotically improved in the worst case. Specifically, we show a family of loss functions $g : [n] \to [m]$ for which the support size of each X_i is at least $\binom{m}{m-n+i}$. For proving this result, it is convenient to use the following notion of a *states graph* of a Markov chain.

Definition 6. (States graph) Let $\bar{X} = (X_0, \ldots, X_n)$ be a Markov chain, and let V_i denote the support set of X_i. Assume, without loss of generality, that the sets V_i are pairwise disjoint. The *states graph* of $\bar{X}$ is a weighted directed graph (G, f) where

- $G = (V, E)$ is a layered graph in which V_i is the set of i-th level nodes and E contains the edges (u, v) such that, for some i, we have $u \in V_i$, $v \in V_{i+1}$ and v is in the support of X_{i+1} conditioned on $X_i = u$.
- For any $u \in V_i$ and $v \in V_{i+1}$, we have $f(u, v) = \Pr[X_{i+1} = v | X_i = u]$.

An example for a states graph of a simple lossy chain appears in Figure 3.

We define, for each node $v \in V \backslash V_0$, a probabilities vector which contains the probability of each starting value given that X_j was chosen to be v.

Definition 7. (Probabilities vector) Let $g : [n] \to [m]$ be a loss function such that $g(n) = m$. Let $\bar{X} = (X_0, \ldots, X_n)$ be a lossy chain realizing g with $|\mathsf{supp}(X_0)| = m$, and let G be its states graph. Let $v \in V_j$ be a node in layer j of G. The *probabilities vector* of v is a vector $\bar{v} \in \mathbb{R}^m$ such that $\bar{v}[i] = \Pr[X_0 = e_i | X_j = v]$, where e_i is the element with index i in V_0, and $\bar{v}[i]$ is the ith coordinate of $\bar{v}$. We say that a vector $\bar{u} \in \mathbb{R}^m$ *fits* layer j of G if $\bar{u}$ has $g(j)$ entries of value $\frac{1}{g(j)}$ and the other entries are 0.

Note that if $\bar{v}$ is the probabilities vector of a node $v \in V_j$, then $\bar{v}$ necessarily fits layer j. However, the converse is not necessarily true.

Our negative result relies on the fact that the probabilities vector of any node in the states graph is a convex linear combination of the probabilities vectors of its parents (that is, a linear combination with positive coefficients that add up to 1).

Lemma 3. *Let $\bar{X} = (X_0 \ldots, X_n)$ be a lossy chain with states graph $G = (V, E)$. For any $1 \leq j \leq n$, let $v \in V_j$ be a node of G and $u_1, \ldots, u_k \in V_{j-1}$ be all the nodes such that $(u_i, v) \in E$. Then $\bar{v}$, the probabilities vector of v, is a convex linear combination of $\bar{u}_1, \ldots, \bar{u}_k$, the probabilities vectors of $u_1, \ldots, u_k$.*

Proof. We will show that the following holds:

$$\bar{v} = \sum_{i=1}^{k} \bar{u}_i \cdot \frac{f(u_i, v)}{\Pr[X_j = v]}.$$

For $e \in [m]$, let $\bar{v}[e]$ be the coordinate of $\bar{v}$ with index e, and let $x_e \in V_0$ be the starting value it corresponds to. By the definitions of the states graph and probabilities vectors we get:

$$\begin{aligned}
\bar{v}[e] &= \Pr[X_0 = x_e | X_j = v] \\
&= \sum_{i=1}^{k} \Pr[X_0 = x_e | X_j = v, X_{j-1} = u_i] \cdot \Pr[X_{j-1} = u_i | X_j = v] \\
&= \sum_{i=1}^{k} \Pr[X_0 = x_e | X_j = v, X_{j-1} = u_i] \cdot \frac{\Pr[X_j = v | X_{j-1} = u_i] \cdot \Pr[X_{j-1} = u_i]}{\Pr[X_j = v]} \\
&= \sum_{i=1}^{k} \Pr[X_0 = x_e | X_{j-1} = u_i] \cdot \frac{\Pr[X_j = v | X_{j-1} = u_i] \cdot \Pr[X_{j-1} = u_i]}{\Pr[X_j = v]} \\
&= \sum_{i=1}^{k} \Pr[X_0 = x_e | X_{j-1} = u_i] \cdot \frac{f(u_i, v) \cdot \Pr[X_{j-1} = u_i]}{\Pr[X_j = v]} \\
&= \sum_{i=1}^{k} \bar{u}_i[e] \cdot \frac{f(u_i, v) \cdot \Pr[X_{j-1} = u_i]}{\Pr[X_j = v]}.
\end{aligned}$$

This implies that $\bar{v}$ can be expressed as a convex linear combination of $\bar{u}_1, \ldots, \bar{u}_k$. □

The main theorem of this section shows a tight negative result on the efficiency of a lossy chain realizing a concrete family of loss functions.

Theorem 2. *Let m, n be positive integers such that $m \geq n$ and let $g_{m,n} : [n] \to [m]$ be the loss function defined by $g_{m,n}(i) = m - n + i$. Let $(X_0, \ldots, X_n)$ be a lossy chain realizing $g_{m,n}$. Then, for any $0 < i \leq n$, it holds that $|\mathsf{supp}(X_i)| \geq \binom{m}{m-n+i}$.*

The theorem relies on the following technical lemma, which will imply a lower bound on the number of probabilities vectors from level i required to span a probabilities vector from level $i+1$.

Lemma 4. *Let $\bar{v} \in \mathbb{R}^n$ be a 0-1 vector of Hamming weight k. Let $\bar{u}_1, \ldots, \bar{u}_m$ be 0-1 vectors of Hamming weight $k-1$. If $\bar{v}$ is a linear combination of $\bar{u}_1, \ldots, \bar{u}_m$ with positive coefficients, then $m \geq k$.*

Proof. Let $\bar{v}$ and $\bar{u}_1, \ldots, \bar{u}_m$ be as above. We write $\bar{v} = \sum_{j=1}^{m} a_j \bar{u}_j$ where $a_j > 0$ for all j. Assume towards a contradiction that $m < k$. Since the support set of each $\bar{u}_j$ is contained in that of $\bar{v}$, this implies that there exists a coordinate h such that $\bar{v}_h = 1$ and $\bar{u}_j[h] = 1$ for all $1 \leq j \leq m$. Thus we have $1 = \sum_{j=1}^{m} a_j$. Moreover, there is an index $h' \neq h$ for which $\bar{v}_{h'} = 1$ and $\bar{u}_j[h'] = 0$ for some $1 \leq j \leq m$. This implies that $1 = \sum_{j=1}^{m} a_j - a_{h'}$. Combining the two equalities, we get that $a_{h'} = 0$, contradicting the assumption that $a_j > 0$ for all j. □

We are now ready to prove Theorem 2.

Proof. Let $\bar{X} = (X_0, \ldots, X_n)$ be a lossy chain realizing $g_{m,n}$. By Claim 1, we may assume without loss of generality that X_0 is uniform over a set of size $g_{m,n}(n) = m$ and X_n has support of size 1.

Let $V_0, \ldots, V_n$ be the layers in the states graph of $\bar{X}$. We prove by induction that, for any $i \in [n]$ and for any of the $\binom{m}{m-n+i}$ probabilities vectors $\bar{v}$ which fit layer i, there is a node $v \in V_i$ such that $\bar{v}$ is the probabilities vector of v. The base case is $i = n$. In this case, the probabilities vector of the (single) node in V_n is $(1/m, \ldots, 1/m)$, which is the only vector which fits level n.

We now assume that the claim holds for layer $i+1$ and prove it for layer i. Let $\bar{u}$ be a vector which fits layer i. By the induction hypothesis, we know that for any vector $\bar{v}$ which fits layer $i+1$ there is a corresponding node $v \in V_{i+1}$. Let $v \in V_{i+1}$ be such a node for which the support set of $\bar{v}$ contains that of $\bar{u}$. By Lemma 3, $\bar{v}$ is a convex linear combination of the probability vectors of its parents u_i. Note that each probabilities vector of a parent node u_i is a scalar multiple of a 0-1 vector of weight $g_{m,n}(i)$ whereas $\bar{v}$ is a scalar multiple of a 0-1 vector of weight $g_{m,n}(i+1)$. By Lemma 4 and the fact that $g_{m,n}(i+1) = g_{m,n}(i) + 1$, the probability vectors $\bar{u}_i$ of the parent nodes u_i have support sets that cover *all* $m - n + i + 1$ subsets of size $m - n + i$ of the support set of $\bar{v}$. In particular, one of the $\bar{u}_i$ must coincide with $\bar{u}$. Since the above holds for any $\bar{u}$ which fits layer i, this concludes the proof of the claim and the theorem. □

Corollary 1. *For every $\varepsilon > 0$ there is an infinite family of loss functions $g_n : [n] \to [m(n)]$, where $m(n) = \lceil n^{1+\varepsilon} \rceil$, such that the information rate of any lossy chain realizing g_n is $O\left(\frac{1}{n}\right)$.*

Proof. We define $g_n : [n] \to [m(n)]$ such that, for any $i \in [n]$, $g_n(i) = m(n) - n + i$. Using Theorem 2, for any $0 < i < n$ the support size of X_i is at least $\binom{m}{m-n+i}$. Therefore, letting $m = m(n)$, we have

$$\rho_n = \min_{0 \leq i \leq n} \frac{\log g_n(n)}{\log |\mathsf{supp}(X_i)|} \leq \frac{\log m}{\log |\mathsf{supp}(X_1)|} \leq \frac{\log m}{\log \binom{m}{m-n+1}} = \frac{\log m}{\log \binom{m}{n-1}}.$$

Using the bound $\left(\frac{a}{b}\right)^b \leq \binom{a}{b}$ and the fact that $m \geq n^{1+\varepsilon}$ we get:

$$\frac{\log m}{\log \binom{m}{n-1}} \leq \frac{\log m}{(n-1) \cdot \log\left(\frac{m}{n-1}\right)} \leq \frac{1}{n-1} \cdot \frac{\log m}{\log\left(n^{\varepsilon}\right)} = \frac{1}{n-1} \cdot \frac{1+\varepsilon}{\varepsilon} = O(1/n)$$

as required. □

4 Fractional Secret Sharing

In this section, we define the notion of *fractional secret sharing* and show how to realize it via the use of lossy chains.

4.1 Definitions

An instance of the fractional secret sharing problem is specified by a *fractional access structure*. Recall that a traditional access structure specifies which subsets of parties can reconstruct the secret, where the remaining sets of parties should learn nothing about the secret. A fractional access structure generalizes this by allowing full control on the amount of information learned by each set of parties.

Definition 8. (Fractional access structure) Let $P = \{p_1, \ldots, p_n\}$ be a finite set of parties and let m be an integer. A function $f : 2^P \to [m]$ is *monotone* if $B \subseteq C$ implies that $f(B) \geq f(C)$. A *fractional access structure* is a monotone function $f : 2^P \to [m]$, with $f(\emptyset) = m$. We say that f is *symmetric* if $f(B)$ depends only on $|B|$.

Note that if we limit the range of f to $\{1, m\}$ then f corresponds to a traditional access structure. We now formally define our notion of fractional secret sharing.

Definition 9. (Fractional secret sharing scheme) Let $f : 2^P \to [m]$ be a fractional access structure and let S be a finite secret-domain. Let D be a randomized algorithm which outputs a uniformly random $s \in S$ together with an n-tuple of shares $(s_1, \ldots, s_n)$. We say that D is a *fractional secret-sharing scheme realizing f with secret-domain S* if there exists a positive integer k such that the following holds: For every $Q \subseteq P$, and any possible share vector s_Q of parties in Q, the distribution of s conditioned on the event that parties in Q receive the shares s_Q is uniform over a subset of S of size $(f(Q) - 1) \cdot k + 1$. If the above holds with $k = 1$, we say that D *strictly realizes* f.

Note that our default notion of realizing a fractional access structure views the structure as only specifying a kind of *ratio* between the amount of uncertainty of different sets, without specifying the absolute amount of uncertainty or the size of the secret-domain. This relaxation is needed in order to capture standard access structures as a special case. Also note that the above definition generates a random secret along with the shares, unlike most traditional definitions of secret sharing which do not refer to any particular distribution over the secret domain. As in the case of traditional secret sharing, we measure the complexity by comparing the size of the biggest share-domain to the size of the secret-domain.

4.2 Fractional Secret Sharing from Lossy Chains

We now apply the positive results from Section 3.2 towards realizing any fractional access structure.

Theorem 3. *For any fractional access structure $f : 2^P \to [m]$, there exists a fractional secret sharing scheme which strictly realizes f.*

Proof. Without loss of generality, assume that $f(P) = 1$ and $f(\emptyset) = m$. We shall use $S = [m]$ as the secret-domain. Let $\alpha_0, ..., \alpha_l$ be all the different values in the range of f in increasing order; that is, $\alpha_0 < ... < \alpha_l$. By our assumptions, we have $\alpha_0 = 1$ and $\alpha_l = m$. Define a loss function $g : [l] \to [m]$ such that $g(i) = \alpha_i$ and let $\bar{X}$ be a lossy chain realizing g. The share generation algorithm D can now proceed as follows:

1. Sample values $(x_0, \ldots, x_l)$ from $\bar{X}$ and let $s = x_0$;
2. For every subset of parties $Q \subseteq P$, let $f(Q) = \alpha_j$. Use a traditional $|Q|$-out-of-$|Q|$ secret sharing scheme to share x_j into $s_{Q,1}, \ldots, s_{Q,|Q|}$ (e.g., using additive secret sharing) and give the j-th party in Q the share $s_{Q,j}$.

We now show that D is a fractional secret sharing scheme strictly realizing f. Let $Q \subseteq P$ be a subset of parties. By the properties of the underlying $|Q|$-out-of-$|Q|$ scheme, the information available to parties in Q is equivalent to learning all values x_j such that $f(Q') = \alpha_j$ for some $Q' \subseteq Q$. By the monotonicity of f this means the parties in Q learn x_i, where i is the index such that $f(Q) = \alpha_i$, and possibly additional values x_j for $j > i$. By the Markov property of a lossy chain, the distribution of the secret s conditioned on the above values x_i and x_j is uniform over a set of size $g(i) = \alpha_i = f(Q)$, as required. □

We remark that if $f(P) \neq 0$, we can add another party p' to the set of parties and set $f(Q)$ to 0 for every subset Q containing p'. We can then execute the proposed algorithm and "throw away" all the shares of p'.

Similarly to traditional secret sharing, the size of the shares produced by the above algorithm can be exponential in the number of parties. This can be avoided in the case of *symmetric* fractional access structures.

Theorem 4. *Let $f : 2^P \to [m]$ be a symmetric fractional access structure with $f(\emptyset) = m$. Then there exists a fractional secret sharing scheme D which (strictly) realizes f with secret-domain $[m]$, where the bit-length of each share is at most $n \cdot \lceil \log \max\{n, m\} \rceil$.*

Proof. As before, let $\alpha_1, \ldots, \alpha_l$ be all the different values in the range of f in increasing order and define $g : [l] \to [m]$ such that $g(i) = \alpha_i$. We now define D as follows:

1. Generate values $\bar{x} = (x_0, \ldots, x_l)$ for the cyclic intervals lossy chain realizing g, and let $s = x_0$. Furthermore, let $a_1, \ldots, a_{l-1}$ be the starting values of the cyclic intervals defining $\bar{x}$ (see Remark 1).
2. For every $i \in [n]$, let α_j be the value such that for any subset of parties $Q \subseteq P$ of size i we have $f(Q) = \alpha_j$. Use Shamir's i-out-of-n threshold secret sharing scheme to create shares of a_j and give one share to each of the parties in P.

We now show that D is a fractional secret sharing scheme. For every subset of parties $Q \subseteq P$, the parties can reconstruct all the values out of $x_0, \ldots, x_n$ that were shared in a threshold scheme requiring $|Q|$ or less parties. This means that if $f(Q) = \alpha_j$, the parties of Q can reconstruct $a_j, \ldots, a_l$. By the definition of the cyclic intervals lossy chain, the parties can reconstruct $x_j, \ldots, x_l$ from $a_j, \ldots, a_l$ and since $x_j, \ldots, x_l$ were generated as values from a lossy chain realizing g we see that the secret s conditioned on $X_j = x_j, \ldots, X_l = x_l$ is distributed uniformly over a set of size α_j, where $\alpha_j = f(Q)$ as required.

We are left with showing that the size of share for each party is no more than $n \cdot \lceil \log(\max\{n, m\}) \rceil$. Each party receives n different shares, one from each invocation of the threshold secret sharing algorithm done by D. The secrets shared are $a_1, \ldots, a_l$ where we recall that all of them are values picked from at most m values. Using Shamir's threshold secret sharing scheme, each of the values is shared with shares of size $\lceil \log(\max\{n, m\}) \rceil$. This amounts to a share size of at most $n \cdot \lceil \log(\max\{n, m\}) \rceil$ for each party, as required. □

5 Conclusions and Open Questions

We introduced the notion of lossy chains – Markov chains which gradually lose information about an initial secret in a controlled fashion. We presented an efficient construction of lossy chains and a matching negative result on the efficiency of lossy chains. Finally, we have shown how lossy chains can be used to realize fractional secret sharing, a natural generalization of traditional secret sharing which supports a fine-grained control over the amount of uncertainty about the secret.

While we essentially settle the main complexity question about lossy chains, it remains open to obtain a characterization of the best achievable information rate for a given loss function g.

The most interesting open question regarding the complexity of fractional secret sharing is to settle the case of *symmetric* fractional access structures, which naturally generalize threshold access structures. While the latter can be realized by an ideal scheme in which the size of each share is equal to the size of the secret (for a sufficiently large secret), we do not know whether an analogous result holds in the fractional domain.

Acknowledgement. We thank Jonathan Yaniv for his contribution to the proof in Appendix A.

References

[1] D. Bertsekas and R. Gallager. Data networks, 1992.

[2] G. Blakley and C. Meadows. Security of ramp schemes. In *Advances in Cryptology*, pages 242–268. Springer, 1985.

[3] G. R. Blakley. Safeguarding cryptographic keys. In *National Computer Conference*, page 313. AFIPS Press., 1979.

[4] R. Cleve. Controlled gradual disclosure schemes for random bits and their applications. In *Advances in Cryptology-CRYPTO'89 Proceedings*, pages 573–588. Springer, 1990.

[5] O. Farràs and C. Padró. Extending brickell-davenport theorem to non-perfect secret sharing schemes. *IACR Cryptology ePrint Archive*, 2012:595, 2012.

[6] Y. Ishai, E. Kushilevitz, and O. Strulovich. Lossy chains and fractional secret sharing. In *Symposium on Theoretical Aspects of Computer Science (STACS)*, 2013.

[7] M. Ito, A. Saito, and T. Nishizeki. Secret sharing scheme realizing general access structure. *Electronics and Communications in Japan*, 72(9):56–64, 1989.

[8] K. Kurosawa, K. Okada, K. Sakano, W. Ogata, and S. Tsujii. Nonperfect secret sharing schemes and matroids. In *Proceedings of EUROCRYPT '93*, pages 126–141, 1993.

[9] A. RRNYI. On measures of entropy and information. In *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pages 547–561, 1961.

[10] A. Shamir. How to share a secret. *Communications of the ACM*, 22(11):612–613, 1979.

[11] C.E. Shannon. Communication Theory of Secrecy Systems. *Journal, vol*, 28(4):656–715, 1949.

[12] H. Yamamoto. Secret sharing system using (k, L, n) threshold scheme. *Electronics and Communications in Japan (Part I: Communications)*, 69(9):46–54, 1986.

A Optimality of Uniform Distributions

The definition of lossy chains requires that the conditional distribution of the initial value X_0 be *uniformly* over a set of a specified size. In this section we justify this restriction by showing that a uniform distribution on secrets minimizes the variance of the number of guesses required for guessing a secret, where minimization is over all distributions (with arbitrary support size) in which the *expected* number of guesses is the same.

In more detail, we consider an abstract game for "guessing" a secret. We show that this game has an optimal strategy that can be used by any rational solver, and then prove that if the required expected number of trials by the player in the game is μ and we wish to minimize the variance of the number of trials, we should pick a uniform distribution with a mean value of μ.

First, we define our abstract game:

Definition 10. (The guessing game) Let $V = \{v_1, \ldots, v_n\}$ be a set and let X be a distribution over V. Let $p_1, \ldots, p_n$ be probabilities such that, for any $i \in [n]$, we have $\Pr[X = v_i] = p_i$. In the *abstract guessing game* of X, a secret value $x \in V$ is chosen according to X, and in round i the player chooses a guess $g_i \in V$. If $g_i = x$, the game ends. Otherwise, it continues to round $i + 1$. The goal of the player is to minimize the number of rounds.

A *strategy* for a player playing the guessing game assigns for every sequence of guesses, $g_1, ..., g_{i-1}$, a probability distribution on the choice of g_i. We want to find the optimal strategy for playing the guessing game. We emphasize that X is known to the player. First, note that a strategy for the guessing game over X can be defined as guessing a *permutation* over V (any strategy which may repeat a guess twice, can be transformed into a strategy that never guesses a value twice and is not inferior to the original strategy, by skipping any guess that has

already been made). Also note that for any randomized strategy, there exists a deterministic strategy with a lower or equal expected number of guesses (where here the expectation is only over the choice of the secret $x \in V$). To see that, take a randomized strategy and consider the different possible permutations on V given the random coins of the player. There exists a permutation π such that π achieves the minimal expected number of guesses. By always choosing π, we get a deterministic strategy which is at least as good as the randomized one.

We now show that the optimal strategy for the game is guessing the values in non-increasing order of probabilities:

Lemma 5. *Let $V = \{v_1, \ldots, v_n\}$ be a set of values and let X be a distribution over V such that for any i, we denote $\Pr[X = v_i] = p_i$. A permutation π over V is the optimal strategy for the guessing game over X, if and only if π is in a non-increasing order of probabilities of all values in V (i.e., for $i < j$ it holds that $p_{\pi(i)} \geq p_{\pi(j)}$).*

Proof. Let $\pi_{OPT} = (v_{i_1}, \ldots, v_{i_n})$ be a permutation that minimizes the expected number of rounds in the guessing game over X. Assume towards a contradiction that there exist two neighboring values in π_{OPT}, v_{i_j} and $v_{i_{j+1}}$ such that $p_{i_j} < p_{i_{j+1}}$. Let π be the permutation which is the result of swapping v_{i_j} and $v_{i_{j+1}}$ in π_{OPT}.

Let Y_{OPT} and Y be the random variables of the number of rounds needed to guess a value chosen according to X using permutations π_{OPT} and π (respectively) as strategies. We now compare the expected number of rounds for each strategy. By dividing into cases, we get:

$$\begin{aligned} E[Y] &= \left(1 - p_{i_j} - p_{i_{j+1}}\right) E\left[Y | X \neq v_{i_j} \wedge X \neq v_{i_{j+1}}\right] \\ &\quad + p_{i_j} \cdot E\left[Y | X = v_{i_j}\right] + p_{i_{j+1}} \cdot E\left[Y | X = v_{i_{j+1}}\right] \\ &= \left(1 - p_{i_j} - p_{i_{j+1}}\right) E\left[Y_{OPT} | X \neq v_{i_j} \wedge X \neq v_{i_{j+1}}\right] + p_{i_j} \cdot (j+1) + p_{i_{j+1}} \cdot j \\ &= \left(1 - p_{i_j} - p_{i_{j+1}}\right) E\left[Y_{OPT} | X \neq v_{i_j} \wedge X \neq v_{i_{j+1}}\right] + \left(p_{i_j} + p_{i_{j+1}}\right) \cdot j + p_{i_j}. \end{aligned}$$

We know that $E\left[Y | X \neq v_{i_j} \wedge X \neq v_{i_{j+1}}\right] = E\left[Y_{OPT} | X \neq v_{i_j} \wedge X \neq v_{i_{j+1}}\right]$ since π_{OPT} and π are identical in such cases. In addition, we can see that for the expected number of rounds for π_{OPT} we have:

$$\begin{aligned} E[Y_{OPT}] &= \left(1 - p_{i_j} - p_{i_{j+1}}\right) E\left[Y_{OPT} | X \neq v_{i_j} \wedge X \neq v_{i_{j+1}}\right] + p_{i_j} \cdot j + p_{i_{j+1}} \cdot (j+1) \\ &= \left(1 - p_{i_j} - p_{i_{j+1}}\right) E\left[Y_{OPT} | X \neq v_{i_j} \wedge X \neq v_{i_{j+1}}\right] + \left(p_{i_j} + p_{i_{j+1}}\right) \cdot j + p_{i_{j+1}}. \end{aligned}$$

Since $p_{i_j} < p_{i_{j+1}}$, we get that $E[Y] < E[Y_{OPT}]$, in contradiction to π_{OPT} being an optimal strategy.

Therefore, we know that a strategy is optimal, only if the values appear in the permutation in a non-increasing order. □

Since there exists an optimal strategy for the game, we can assume that any rational solver will use it. We wish to select the secret value from a distribution that allows us to control the expected number of guesses the player will need. In addition, we also want to minimize the variance under this condition. We now show that the optimal distribution for this purpose is a uniform distribution over a minimal set of elements with which the required expected number of guesses can be achieved.

First, since we know the optimal strategy, we can simplify our notation: Let n be a big enough integer, and X a distribution over $[n]$ such that for any $i \in [n]$ we have $\Pr[X = i] \geq$

$\Pr[X = i+1]$. In this case, X is the random variable whose value is also the number of guesses needed by an optimal player for each secret. We now wish to find such a distribution X such that $E[X] = \mu$, for a given integer μ, where $V[X]$, the variance of X, is minimal subject to the restriction on $E[X]$.

Theorem 5. *Let $\mu, n \in \mathbb{N}$ be such that $2\mu < n$ and let $p_1, \ldots, p_n$ be the following values:*

$$p_i = \begin{cases} \frac{1}{2\mu-1} & i \leq 2\mu - 1 \\ 0 & \textit{otherwise.} \end{cases}$$

Then, the distribution X defined by $\Pr[X = i] = p_i$, has a mean value of μ and, for every other monotone non-increasing distribution Y over $[n]$, it holds that $V[X] \leq V[Y]$.

We prove this theorem in two steps. The first step will be to show that for any distribution Y over $[n]$ with an expected value of μ and minimal variance, the support of Y is a subset of $[2\mu - 1]$. The second step will be to show that if a distribution Y over $[n]$ has a mean value of μ and $\mathsf{supp}(Y) \subseteq [2\mu - 1]$, then for any $i \in [n]$ we have $\Pr[Y = i] = p_i$, and hence $Y = X$.

Lemma 6. *Let $m, n \in \mathbb{N}$ be two integers, and let X be a monotone non-increasing distribution over $[n]$ with a mean value of μ. Let k be the maximal index such that $Pr[X = k] > 0$. If $\mathsf{supp}(X) \nsubseteq [2\mu - 1]$, then there exists $i \in [k-2]$ such that $Pr[X = i] > Pr[X = i+1] > 0$.*

Proof. Let k be the maximal index such that $\Pr[X = k] > 0$. Assume towards contradiction that for any $i \in [k-2]$, we have $\Pr[X = i] = \Pr[X = i+1]$ (we know that X is non-increasing). Denote $\Pr[X = i] = p$, and $\Pr[X = k] = p'$. We can see that $E[X] = p \cdot \sum_{i=1}^{k-1} i + p' \cdot k$. Therefore, we can tell that $E[X] = p \cdot k \cdot \frac{k-1}{2} + p' \cdot k$. In addition, we know that $1 = \sum_{i=1}^{k-1} p + p' = (k-1)p + p'$. By joining these we get:

$$\begin{aligned} E[X] &= p \cdot k \cdot \frac{k-1}{2} + p' \cdot k \\ &= k \cdot \left(\frac{p \cdot (k-1) + p'}{2} + \frac{p'}{2} \right) \\ &= k \cdot \frac{1+p'}{2}. \end{aligned}$$

Finally, we note that $E[X] = \mu$, and $k > 2\mu - 1$, which means $k \geq 2\mu$. Therefore:

$$\mu = k \cdot \frac{1+p'}{2} \geq \mu\left(1+p'\right).$$

This contradicts the fact that $p' > 0$. □

Lemma 7. *Let X be monotone non-increasing distribution over $[n]$ with a mean value of μ and minimal variance, then $\mathsf{supp}(X) \subseteq \{1, \ldots, 2\mu - 1\}$.*

Proof. Assume towards a contradiction that $\mathsf{supp}(X) \nsubseteq \{1, \ldots, 2\mu - 1\}$, and let k be the maximal index such that $p_k > 0$. By Lemma 6, there exists an index $i < k - 1$ such that $p_i > p_{i+1}$. Let $\varepsilon_1 > 0$ and $\varepsilon_2 = (k - i - 1)\varepsilon_1$ be values such that:

$$\begin{aligned} p_k &\geq \varepsilon_1 \\ p_i - p_{i+1} &\geq 2\varepsilon_2 + \varepsilon_1. \end{aligned}$$

We modify X to construct a new distribution Y with probabilities $p'_1, \ldots, p'_n$:

$$p'_l = \begin{cases} p_i - \varepsilon_2 & l = i \\ p_{i+1} + \varepsilon_1 + \varepsilon_2 & l = i+1 \\ p_k - \varepsilon_1 & l = k \\ p_l & \text{otherwise.} \end{cases}$$

Note that for any $l \in [n-1]$ we have $p'_l \geq p'_{l+1}$. Also note that Y is indeed a valid distribution:

$$\sum_{l=1}^{n} p'_l = \sum_{l=1}^{n} p_l - \varepsilon_2 + (\varepsilon_1 + \varepsilon_2) - \varepsilon_1 = 1$$

We calculate the mean value of Y:

$$\begin{aligned} E[Y] &= \sum_{l=1}^{n} l \cdot p'_l \\ &= \sum_{l=1}^{n} l \cdot p_l - i \cdot \varepsilon_2 + (i+1) \cdot (\varepsilon_1 + \varepsilon_2) - k \cdot \varepsilon_1 \\ &= \mu + \varepsilon_2 + (i+1-k)\,\varepsilon_1 \\ &= \mu + (k-i-1) \cdot \varepsilon_1 + (i+1-k)\,\varepsilon_1 \\ &= \mu. \end{aligned}$$

Therefore Y is a valid distribution with a mean value of μ. Lastly, we find its variance:

$$\begin{aligned} V[Y] &= E\left[Y^2\right] - (E[Y])^2 \\ &= \sum_{l=1}^{n} l^2 \cdot p'_l - \mu^2 \\ &= \sum_{l=1}^{n} l^2 \cdot p_l - i^2\varepsilon_2 + (i+1)^2 (\varepsilon_1 + \varepsilon_2) - k^2\varepsilon_1 - \mu^2 \\ &= V[X] - i^2\varepsilon_2 + (i+1)^2 (\varepsilon_1 + \varepsilon_2) - k^2\varepsilon_1. \end{aligned}$$

We are left with evaluating $V[Y] - V[X]$:

$$\begin{aligned} V[Y] - V[X] &= -i^2\varepsilon_2 + (i+1)^2 (\varepsilon_1 + \varepsilon_2) - k^2\varepsilon_1 \\ &= \varepsilon_1 \left((i+1)^2 - k^2\right) + \varepsilon_2 \left((i+1)^2 - i^2\right) \\ &= \varepsilon_1 (i+1-k)(i+1+k) + \varepsilon_1 (k-i-1)(2i+1) \\ &= \varepsilon_1 (i+1-k)(i+1+k-2i-1) \\ &= \varepsilon_1 (i+1-k)(k-i). \end{aligned}$$

Since $i < k-1$, we get $V[Y] - V[X] < 0$. In contradiction to X having minimal variance under the required conditions. □

We are now left with the final step: showing that if the support of X is $[2\mu-1]$ then there is only one possible distribution.

Lemma 8. *Let X be a non-increasing distribution over $[n]$ with a mean value of μ and $\mathsf{supp}(X) \subseteq [2\mu-1]$. Then:*

$$\Pr[X=i] = \begin{cases} \frac{1}{2\mu-1} & i \leq 2\mu-1 \\ 0 & \textit{otherwise.} \end{cases}$$

Proof. Let X be the non-increasing distribution over $[n]$ such that for any $i \leq 2\mu-1$ we have $\Pr[X=i] = \frac{1}{2\mu-1}$ and 0 otherwise. First, we note that $E[X] = \mu$ since:

$$\begin{aligned} E[X] &= \sum_{i=1}^{n} i \cdot \Pr[X=i] \\ &= \sum_{i=1}^{2\mu-1} i \cdot \Pr[X=i] \\ &= \frac{1}{2\mu-1} \cdot \sum_{i=1}^{2\mu-1} i \\ &= \frac{1}{2\mu-1} \cdot (2\mu-1) \cdot \frac{1+(2\mu-1)}{2} \\ &= \mu. \end{aligned}$$

Next, we show that for any non-increasing distribution Z over $[n]$ such that $\mathsf{supp}(Z) \subseteq [2\mu-1]$, we have $E[Z] \leq \mu$.

Let Y be a non-increasing distribution over $[n]$ such $\mathsf{supp}(Y) \subseteq [2\mu-1]$ and $E[Y]$ is maximal under these conditions. Assume towards a contradiction, that $Y \neq X$, then there exists an index $i \in [2\mu-2]$ and $\varepsilon > 0$ such that $\Pr[Y=i] - \Pr[Y=i+1] = 2\varepsilon$. We define Y' as follows:

$$\Pr[Y'=j] = \begin{cases} \Pr[Y=i] - \varepsilon & j = i \\ \Pr[Y=i+1] + \varepsilon & j = i+1 \\ \Pr[Y=j] & \text{otherwise.} \end{cases}$$

We can see that Y' is still a monotone non-increasing distribution where $\mathsf{supp}(Y') \subseteq [2\mu-1]$ and that $E[Y'] = \mu + \varepsilon$. In contradiction to Y having the maximal expectation under these conditions. Therefore, $Y = X$. Since $E[X] = \mu$, we can see for every non-increasing distribution Z over $[n]$ such that $\mathsf{supp}(Z) \subseteq [2\mu-1]$, we have $E[Z] \leq \mu$.

This means that $E[X]$ is the maximal mean value and since $E[X] = \mu$ we get that X is the only distribution satisfying all of the conditions. □

By combining Lemma 7 and Lemma 8, we get a proof for Theorem 5 and can indeed see that the minimal variance for the guessing game with a mean value of μ guesses is received when choosing the secret value from a uniform distribution.

B Relaxation of Uniform Distributions

In this section we consider a relaxation of lossy chains referred to as *ε-close lossy chains*. In an ε-close lossy chain, we substitute the requirement that conditional distributions be uniform with being statistically close to uniform. We show two results for this case. The first result shows how to construct an ε-close lossy chain for any loss function such that the information rate of the chain is $\frac{\log \frac{1}{1-\varepsilon}}{\log m}$. The second result shows that for a sufficiently small ε the upper bound from Section 3.3 still holds. That is, the information rate of an ε-lossy chain is still $O\left(\frac{1}{n}\right)$.

We start with the formal definitions for the distance between two distributions, and ε-close lossy chains.

Definition 11. (Statistical Distance) Let X and Y be two random variables distributed over a set S. We denote the *statistical distance between X and Y* as $d(X,Y)$ which is defined as:

$$d(X,Y) = \frac{1}{2}\sum_{s\in S} |\Pr[X=s] - \Pr[Y=s]|.$$

In addition, we say that a random variable X distributed over S is *ε-close to a uniform distribution of size n for $\varepsilon > 0$* if there exists a random variable Y which is distributed uniformly over a subset of S of size n such that $d(X,Y) \leq \varepsilon$.

We also denote by d the distance between two vectors $\bar{v}$,$\bar{u}$, which we define as:

$$d(\bar{v},\bar{u}) = \frac{1}{2}\sum_{i} |\bar{v}[i] - \bar{u}[i]|.$$

This means that if $\bar{v}$ and $\bar{u}$ are probabilities vectors for distributions X and Y, then $d(\bar{v},\bar{u}) = d(X,Y)$.

Definition 12. (ε-Close Lossy chain) Let $g:[n]\to[m]$ be a loss function, let $\varepsilon > 0$ be a positive constant, and let $\bar{X} = (X_0, X_1, \ldots, X_n)$ be a sequence of random variables. We say that $\bar{X}$ is a *ε-close lossy chain realizing g* if the following conditions hold:

1. $\bar{X}$ is a Markov chain, and
2. for every $i \in [n]$ and every x_i in the support of X_i, the distribution of X_0 conditioned on $X_i = x_i$ is ε-close to a uniform distribution over a set of size $g(i)$.

B.1 Constructing an ε-Lossy Chain

First, as a warm up, consider the case in which $m = \{0,1\}^n$, and we wish to construct a $\frac{1}{2}$-lossy chain realizing a loss function $g:[n]\to[m]$. Let $\bar{Y}$ be the naive lossy chain in which Y_i contains the first $n-i$ bits of the starting value Y_0. We can get our $\frac{1}{2}$-lossy chain $\bar{X}$ by setting X_i to be Y_j where j is the minimal index such that $2^{n-j} \geq f(i)$. Intuitively, this means that the set of possible values for X_i is at most twice than the required size by $g(i)$, which means X_i is a distribution with a statistical distance of at most $\frac{1}{2}$ from a uniform distribution over $g(i)$ values.

This simple construction, in which we "round-up" the values of g to match our existing lossy chain and get an ε-lossy chain will serve as the basis for how we build an ε-lossy chain for

every loss function and every $\varepsilon > 0$. We now show formally how to construct such an ε-lossy chain, and find its information rate.

We start by proving a lemma on the distance between two uniform distributions:

Lemma 9. *Let n and m be two integers, and U_n and U_m be the uniform distribution over $[n]$ and $[m]$ respectively. For $0 < \varepsilon < 1$, if $n \leq m \leq \frac{1}{1-\varepsilon} \cdot n$, then $d(U_n, U_m) \leq \varepsilon$.*

Proof. Since $n \leq m$, the statistical distance between U_n and U_m can be computed as follows:

$$d(U_n, U_m) = \frac{1}{2} \cdot \left(n \cdot \left(\frac{1}{n} - \frac{1}{m} \right) + (m - n) \cdot \frac{1}{m} \right) = 1 - \frac{n}{m}.$$

Since $m \leq \frac{1}{1-\varepsilon} \cdot n$, we can see that $1 - \frac{n}{m} \leq \varepsilon$, as required. □

We now proceed to the construction of a general ε-lossy chain:

Theorem 6. *Let $0 < \varepsilon < 1$, and let $g : [n] \to [m]$ be a loss function. Then there exists an ε-lossy chain realizing g with an information rate of at least $\frac{\log \frac{1}{1-\varepsilon}}{\log m}$.*

Proof. Let $g : [n] \to [m]$ be a loss function and $0 < \varepsilon < 1$. Let k be the minimal integer such that $\lceil \left(\frac{1}{1-\varepsilon} \right)^k \rceil \geq m$. We define the loss function $h : [k] \to [m]$ as:

$$h(i) = \begin{cases} m & i = k \\ \lceil \left(\frac{1}{1-\varepsilon} \right)^i \rceil & i < k. \end{cases}$$

Let $\bar{Y} = (Y_1, \ldots, Y_k)$ be a lossy chain realizing h using the cyclic-intervals consturction shown in Section 3.2. We define our ε-lossy chain $\bar{X} = (X_1, \ldots, X_n)$ as follows: For each $i \in [n]$ set X_i to be Y_j, where j is the minimal integer such that $h(j) \geq g(i)$. $\bar{X}$ is a Markov chain, since for each two variables in it X_k and X_{k+1}, there exists two integers k_1, k_2 such that $X_k = Y_{k_1}$ and $X_{k+1} = Y_{k_2}$, where $k_2 \geq k_1$, and $\bar{Y}$ is a chain in which Y_{k_2} can be determinstically computed from Y_{k_1}. We now turn to showing that $\bar{X}$ is indeed an ε-lossy chain realizing g.

Let i and j be two indexes such that $X_i = Y_j$ in our consturction. By the fact that j is the minimal index such that $g(i) \leq h(j)$, we get:

$$h(j-1) < g(i) \leq h(j).$$

Since $g(i)$ is an integer, we also know that:

$$h(j-1) + 1 \leq g(i) \leq h(j).$$

In addition, by the definition of h we can bound the the ratio between $h(j)$ and $h(j-1)$:

$$\frac{h(j)}{h(j-1)+1} \leq \frac{\lceil \left(\frac{1}{1-\varepsilon} \right)^j \rceil}{\lceil \left(\frac{1}{1-\varepsilon} \right)^{j-1} \rceil + 1} \leq \frac{\left(\frac{1}{1-\varepsilon} \right)^j + 1}{\left(\frac{1}{1-\varepsilon} \right)^{j-1} + 1} \leq \frac{\left(\frac{1}{1-\varepsilon} \right)^j}{\left(\frac{1}{1-\varepsilon} \right)^{j-1}} = \frac{1}{1-\varepsilon}.$$

We can now apply Lemma 9, and deduce that the statistical distance of X_i from a uniform distribution over $g(i)$ is less than ε as required.

We are left with showing that the information rate of $\bar{X}$ is indeed $\frac{\log \frac{1}{1-\varepsilon}}{\log m}$. By Theorem 1, we know that the information rate of $\bar{Y}$ is at least $\frac{1}{k-1}$. In addition, we know that k is the minimal integer such that $\lceil \left(\frac{1}{1-\varepsilon}\right)^k \rceil \geq m$ and therefore $\lceil \left(\frac{1}{1-\varepsilon}\right)^{k-1} \rceil < m$, which means $\left(\frac{1}{1-\varepsilon}\right)^{k-1} < m$. From this we get:

$$k - 1 \leq \log_{\frac{1}{1-\varepsilon}} m.$$

We conclude that the information rate of $\bar{X}$ is at least $\frac{\log \frac{1}{1-\varepsilon}}{\log m}$ by applying a change of the logarithm base. □

B.2 An Upper Bound for ε-Lossy Chains

We now show that for a sufficiently small ε the information rate of an ε-lossy chain is still $O\left(\frac{1}{n}\right)$, where n is the length of the chain.

We use the same definition of a states graph of a lossy chain used in Section 3.2. In particular, for a states graph G of an ε-lossy chain realizing a loss function g, we still say that a vector $\bar{u} \in \mathbb{R}^m$ *fits* layer j of G if $\bar{u}$ has $g(j)$ entries of value $\frac{1}{g(j)}$ and the other entries are 0. For our case we also add the following definition:

Definition 13. (ε-Fitting Vector) Let $\bar{X}$ be an ε-lossy chain realizing a loss function $g : [n] \to [m]$, and let G be its states graph. Let $\bar{u}, \bar{w} \in \mathbb{R}^m$ be two vectors. We say that $\bar{u}$ *ε-fits layer j of G through $\bar{w}$,* if $\bar{w}$ fits layer j of G and $d(\bar{u}, \bar{w}) < \varepsilon$.

We now define a generalization of Theorem 2 for ε-close lossy chains.

Theorem 7. *Let m, n be positive integers such that $m \geq n$ and let $g_{m,n} : [n] \to [m]$ be the loss function defined by $g_{m,n}(i) = m - n + i$. Let $(X_0, \ldots, X_n)$ be an ε-close lossy chain realizing $g_{m,n}$. If $\varepsilon < \frac{1}{32m^3}$ then, for any $0 < i \leq n$, it holds that $|\mathsf{supp}(X_i)| \geq \binom{m}{m-n+i}$.*

As in the proof of Theorem 2, we start with a similar lemma to Lemma 4, which we prove in two steps.

Lemma 10. *Let $\bar{v}, \bar{u}_1, \ldots, \bar{u}_m \in \mathbb{R}^n$ be 0-1 vectors, such that the Hamming weight of $\bar{v}$ is $k \geq 2$, and the Hamming weight of $\bar{u}_1, \ldots, \bar{u}_m$ is $k - 1$. If there exists $0 \leq \varepsilon < \frac{1}{8k+4}$ such that there exists a linear combination of $\bar{u}_1, \ldots, \bar{u}_m$ with non-negative coefficients, denoted $\bar{v}^*$ and $d(\bar{v}, \bar{v}^*) < \varepsilon$, then $m \geq k$.*

Proof. Let $\bar{v}, \bar{u}_1, \ldots, \bar{u}_m \in \mathbb{R}^n$ be such vectors, that is, there exists $0 \leq \varepsilon < \frac{1}{8k+4}$, and a linear combination with non-negative coefficients of $\bar{u}_1, \ldots \bar{u}_m$which we denote by $\bar{v}^*$ such that $d(\bar{v}, \bar{v}^*) < \varepsilon$.

Let $a_1, \ldots, a_m$ be the non-negative coefficients such that $\bar{v}^* = \sum_{i=1}^m a_i \cdot \bar{u}_i$.

Since $d(\bar{v}, \bar{v}^*) < \varepsilon$, we know that for every index $i \in [n]$, the distance between the entries $\bar{v}[i]$ and $\bar{v}^*[i]$ is at most 2ε, which means:

$$1 - 2\varepsilon = \bar{v}[i] - 2\varepsilon < \bar{v}^*[i] < \bar{v}[i] + 2\varepsilon = 1 + 2\varepsilon \tag{1}$$

We denote by U' the subset of $\bar{u}_1, \ldots, \bar{u}_m$ such that for each $\bar{u} \in U'$, $\mathsf{supp}(\bar{u}) \subseteq \mathsf{supp}(\bar{v})$. Because $d(\bar{v}, \bar{v}^*) < \varepsilon$ we get an upper bound on the coefficients of the vectors not in U':

$$\sum_{u_i \notin U'} a_i < 2\varepsilon. \tag{2}$$

Assume towards contradiction that $m < k$. By the pigeonhole principle, there's an index $b \in [n]$ such that $\bar{v}[b] = 1$, and for every $\bar{u_i} \in U'$, $\bar{u}_i[b] = 1$. This allows us to bound $\sum_{u_i \in U'} a_i$. Splitting the vectors to U', and vectors not in U', we get:

$$\bar{v}^*[b] = \sum_{u_i \in U'} a_i \cdot \bar{u}_i[b] + \sum_{u_i \notin U'} a_i \cdot \bar{u}_i[b] = \sum_{u_i \in U'} a_i + \sum_{u_i \notin U'} a_i \cdot \bar{u}_i[b].$$

By combining (1), (2), and the last equation we get:

$$1 - 4\varepsilon < \sum_{u_i \in U'} a_i < 1 + 4\varepsilon. \tag{3}$$

Using this bound, we can see that there exists $j \in [m]$ such that $a_j > \frac{1-4\varepsilon}{|U'|} \geq \frac{1-4\varepsilon}{m}$.

We now move on to find a bound for $\sum_{u_i \in U'} a_i - a_j$: Since the Hamming weight of $\bar{u}_j$ is strictly less than that of $\bar{v}$, there exists an index b' such that $\bar{v}[b'] = 1$ and $\bar{u}_j[b'] = 0$. This gives us an upper bound over $\bar{v}^*[b']$:

$$\bar{v}^*\left[b'\right] = \sum_{u_i \in U'} a_i \cdot \bar{u}_i\left[b'\right] + \sum_{u_i \notin U'} a_i \cdot \bar{u}_i\left[b'\right] < \sum_{u_i \in U'} a_i - a_j + \sum_{u_i \notin U'} a_i \cdot \bar{u}_i\left[b'\right].$$

Again, we combine the last equation with (1) and (2), and get:

$$1 - 4\varepsilon < \sum_{u_i \in U'} a_i - a_j. \tag{4}$$

From (3), (4) and the fact $a_j > 0$, we can see that $a_j < 8\varepsilon$. In addition, we have shown that $a_j > \frac{1-4\varepsilon}{m}$, which shows us that $\frac{1-4\varepsilon}{m} < 8\varepsilon$. Since $\varepsilon < 1$, we can rewrite the last inequality as $\frac{1}{8m+4} < \varepsilon$. Finally, we remind that $\varepsilon < \frac{1}{8k+4}$ and $m < k$, which means $\varepsilon < \frac{1}{8m+4}$, in contradiction to our assumption that $m < k$. Therefore $m \geq k$. □

We now apply Lemma 10 to prove the generalization of Lemma 4 from Section 3.3.

Lemma 11. *Let $\varepsilon > 0$, let $\bar{v}$ be a distribution vector of an ε-close to uniform distribution of size $k \geq 2$, and let $\bar{u}_1, \ldots, \bar{u}_m$ be distribution vectors of ε-close to uniform distributions of size $k-1$ each. If $\bar{v}$ is a convex combination of $\bar{u}_1, \ldots, \bar{u}_m$, $\varepsilon < \frac{1}{16k^3}$, and we denote by $\bar{w}_1, \ldots, \bar{w}_m$ the distribution vectors of all the uniform distributions such that $d(\bar{u}_i, \bar{w}_i) < \varepsilon$, then there are at least k distinct vectors in $\bar{w}_1, \ldots, \bar{w}_m$.*

Proof. Let $\varepsilon > 0$, and let $\bar{v}, \bar{u}, \ldots, \bar{u}_m \in \mathbb{R}^n$ be vectors as mentioned above. Let $\bar{w}_1, \ldots, \bar{w}_m$ be the respective uniform distribution vectors. Since $\bar{v}$ is ε-close to a uniform distribution of size k, there exists a 0-1 vector $\bar{v}^* \in \mathbb{R}^n$, with Hamming weight k, such that $d(\bar{v}, \bar{v}^*) \leq k \cdot \varepsilon$. Similarly, for each $i \in [m]$, $(k-1)\bar{w}_i$ is a 0-1 vector of Hamming weight $k-1$, and it holds that:

$$d((k-1)\bar{w}_i, (k-1)\bar{u}_i) < (k-1)\varepsilon$$

Let $a_1, \ldots, a_m$ be the coefficients such that $\bar{v} = \sum_{i=1}^{m} a_i \bar{u}_i$ where $a_i \geq 0$. We will show that the distance between $\bar{v}^*$ and $\sum_{i=1}^{m} \frac{k}{k-1} a_i \cdot (k-1)\, \bar{w}_i$ is less than $\frac{1}{8k+4}$, thus contradicting Lemma 10.

First, let us bound this distance as the sum of three distances:

$$\begin{aligned} d\left(\bar{v}^*, \sum_{i=1}^{m} \frac{k}{k-1} a_i\,(k-1)\,\bar{w}_i\right) \quad \leq \quad & d\left(\bar{v}^*, k \cdot \bar{v}\right) + d\left(k \cdot \bar{v}, \sum_{i=1}^{m} \frac{k}{k-1} a_i\,(k-1)\,\bar{u}_i\right) \\ & + d\left(\sum_{i=1}^{m} \frac{k}{k-1} a_i\,(k-1)\,\bar{u}_i, \sum_{i=1}^{m} \frac{k}{k-1} a_i\,(k-1)\,\bar{w}_i\right). \end{aligned} \tag{5}$$

Assume toward contradiction that $m < k$.

The first part $d\left(\bar{v}^*, k \cdot \bar{v}\right)$ is less than $k\varepsilon$. The second term can be simplified, and it is in fact zero:

$$d\left(k \cdot \bar{v}, \sum_{i=1}^{m} \frac{k}{k-1} a_i\,(k-1) \cdot \bar{u}_i\right) = d\left(k \cdot \bar{v}, \sum_{i=1}^{m} k \cdot a_i \cdot \bar{u}_i\right) = k \cdot d\left(\bar{v}, \sum_{i=1}^{m} a_i \cdot \bar{u}_i\right) = 0.$$

Lastly, let us simplify the final term:

$$d\left(\sum_{i=1}^{m} \frac{k}{k-1} a_i\,(k-1) \cdot \bar{u}_i, \sum_{i=1}^{m} \frac{k}{k-1} a_i\,(k-1)\,\bar{w}_i\right) \leq k \cdot \sum_{i=1}^{m} a_i \cdot d\left(\bar{u}_i, \bar{w}_i\right) \leq k \cdot \varepsilon \cdot \sum_{i=1}^{m} a_i.$$

We are left with bounding $\sum_{i=1}^{m} a_i$. For every $i \in [m]$, we know that there exists an index j such that $\bar{u}_i[j] > \frac{1}{k-1} - 2\varepsilon$. In addition, $\bar{v}[j] < \frac{1}{k} + 2\varepsilon$. Since $a_i \geq 0$ we get:

$$a_i \cdot \left(\frac{1}{k} - 2\varepsilon\right) < a_i \cdot \left(\frac{1}{k-1} - 2\varepsilon\right) < a_i \bar{u}_i[j] \leq \bar{v}[j] < \frac{1}{k} + 2\varepsilon.$$

Therefore $a_i < \frac{1+2\varepsilon k}{1-2\varepsilon k}$. Since $\varepsilon < \frac{1}{16k^3}$, and $k \geq 2$, we can get a loose bound for a_i: $a_i < 2$. This, combined with $m < k$ gives: $\sum_{i=1}^{m} a_i < 2m < 2k$.

By assigning all the bounds in (5), we get:

$$d\left(\bar{v}^*, \sum_{i=1}^{m} a_i\,(k-1)\,\bar{w}_i\right) \leq \varepsilon k + \varepsilon k \cdot 2k.$$

Since $\varepsilon < \frac{1}{32k^3}$ we get that $\varepsilon k < \frac{1}{32k^2}$. Combined with the fact that $k \geq 2$, we finally arrive at:

$$d\left(\bar{v}^*, \sum_{i=1}^{m} a_i\,(k-1)\,\bar{w}_i\right) \leq \frac{1+2k}{32k^2} < \frac{1}{8k+4}.$$

The last inequality is true since $k \geq 2$.

By Lemma 10, we see that $\sum_{i=1}^{m} a_i\,(k-1)\,\bar{w}_i$ must be a linear combination with more than k distinct vectors as required. □

We can now state and prove the lower bound on the support of the random variables in an ε-close lossy chain.

Theorem 8. *Let m, n be positive integers such that $m \geq n$ and let $g_{m,n} : [n] \to [m]$ be the loss function defined by $g_{m,n}(i) = m - n + i$. Let $(X_0, \ldots, X_n)$ be an ε-close lossy chain realizing $g_{m,n}$. If $\varepsilon < \frac{1}{32m^3}$ then, for any $0 < i \leq n$, it holds that $|\mathsf{supp}(X_i)| \geq \binom{m}{m-n+1}$.*

Proof. Let $\bar{X} = (X_0, \ldots, X_n)$ be an ε-close lossy chain realizing $g_{m,n}$.

Let $V_0, \ldots, V_n$ be the layers in the states graph of $\bar{X}$. We prove by induction that, for any $i \in [n]$ and for any of the $\binom{m}{m-n+i}$ probabilities vectors $\bar{w}$ which fit layer i, there is a node $v \in V_i$ such that $\bar{v}$ ε-fits layer i through $\bar{w}$**.**

The base case is $i = n$. In this case, the probabilities vector of the (single) node v in V_n is ε-close to $(1/m, \ldots, 1/m)$, which is the only vector which fits level n. Therefore $\bar{v}$ ε-fits layer n through $(1/m, \ldots, 1/m)$.

We now assume that the claim holds for layer $i+1$ and prove it for layer i. Let $\bar{w}$ be a vector which fits layer i. By the induction hypothesis, we know that for any vector $\bar{w}'$ which fits layer $i+1$ there is a corresponding node $v \in V_{i+1}$ such that $\bar{v}$ ε-fits layer $i+1$ through $\bar{w}'$.

Let $v \in V_{i+1}$ be a node such that $\bar{v}$ ε-fits layer $i+1$ through $\bar{w}'$ and the support of $\bar{w}'$ contains the support of $\bar{w}$. We note that the conditions of Lemma 11 hold: $\bar{v}$ is a convex combination of the probability vectors of its parents $\bar{u}_1, \ldots, \bar{u}_l$. In addition, $\bar{v}$ is ε-close to a uniform distribution with a support of size $g_{m,n}(i+1)$, and $\bar{u}_1, \ldots, \bar{u}_l$ are all ε-close to uniform distribution of size $g_{m,n}(i)$, where $g_{m,n}(i) = g_{m,n}(i+1) - 1$. Finally, $\varepsilon < \frac{1}{32m^3}$. Therefore by Lemma 11, there exists a parent vector $\bar{u}_j$ which is ε-close to $\bar{w}$, which means that $\bar{u}_j$ ε-fits layer i through $\bar{w}$.

Since there are $\binom{m}{m-n+i}$ different vector which fit layer i, we get that layer i has at least $\binom{m}{m-n+i}$ different nodes, as required. □

C Efficient Generation and Enumeration of Cyclic Intervals

In this section we give a pseudo-code implementation of the lossy chain construction from Section 3.2. We also give an efficient algorithm for enumerating the possible values of the secret X_0 given a value of X_i.

The first function, *generate_lossy_chains* gets two integers x and n, and a loss function $f : [n] \to [m]$, where $f(n) = m$, and returns the cyclic interval's starting points $a_0, \ldots, a_n$ as a list. This means that the first i elements of the result are the value of X_{n-i} in the chain. We note that this version of the algorithm gets the original secret x, and generates the lossy chain with x as its secret. The variable x is maintained through the execution to hold the index of the secret in the last cyclic interval chosen so far.

```
function generate_lossy_chain(int x, int n, function f) {
    list<int> chain;
    for (int i = n - 1; i > 0; i--) {
        int a = random(f(i - 1));
        int b = (x - a) mod f(i);
        chain.add(b);
        x = a;
    }
    return chain;
}
```

The function *random* returns a random number between 0 and its argument (exclusive).

The next algorithm shown, *enumaerate_ cyclic_ intervals*, gets n, an index i, a prefix of the list generated by the previous function, and the loss function f and returns the ith element of the set S_j, that is one of the elements which are the possible value of the original secret given X_j.

```
int enumerate_lossy_chain(int n, function f,
                          list<int> chain, int index) {
    for (int i = chain.size(); i > 0; i--) {
        index = (chain[i - 1] + index) mod f(n - i);
    }
    return index;
}
```

We can see that both functions are efficient, and we conclude that generating the lossy chain for a chosen secret, and recovering the possible values given one of its elements can be done efficiently.