Secure Convertible Codes

Justin Zhang

July 2024

For my grandmother.

Abstract

Large-scale distributed storage systems (DSS) make use of erasure codes to enforce fault tolerance in the event of node failure. Due to observed changing failure rates within these systems, code redundancy tuning, or *code conversion* has been shown to reduce storage cost. Prior work has developed bounds and constructions across many parameters for *convertible codes,* a class of erasure codes optimizing either the access or bandwidth costs of conversion.

In this thesis, we investigate the information-theoretic security of convertible codes under the presence of an eavesdropper whom we enforce to learn nothing of the stored message. While current convertible code constructions are inherently insecure since they are systematic, we present novel constructions that augment existing cost-optimal convertible codes with perfect eavesdropper security. Furthermore, we prove that our constructions maximizes the amount of information that can be stored on such a system, and we give additional constructions and bounds for secure codes when additional information is known about the distribution of eavesdropped nodes.

Acknowledgments

.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Erasure codes are a low storage overhead solution used in large-scale distributed storage systems to apply suitable fault tolerance against node failure [2, 3]. The data to be stored is partitioned into k symbol chunks, where each of these chunks are encoded as n symbols (called a codeword) under a $[n, k]$ erasure code and are distributed onto n nodes in the systems. Maximum distance separable (MDS) codes are often chosen in this setting since they require the minimum amount of storage overhead and guarantee of data integrity in the face of $(n - k)$ node failures. In other words, any k out of n symbols of a codeword can be used to recover the original data.

The parameters n, k are chosen based on the node failure rate, which may change over time. For instance, in periods of high failure rates, n and k may be chosen so that their rate of redundancy $\frac{n}{k}$ is high (at the cost of higher storage overhead), while in periods of low failure rate, n and k may be chosen so that their rate of redundancy is low (at the benefit of lower storage overhead). Kadekodi et al. have shown in prior work that failure rate of disks can vary over time, where significant savings can be made in storage costs by tuning n and k in response [4]. However, re-tuning n and k under the default method of decoding the data under an initial code and re-encoding the data under a new code is costly in terms of I/O, CPU, and network bandwidth resources. This has led to the study of the *code conversion* problem [5, 6, 7, 8]. Code conversion is the process of transforming data encoded under a $[n^I, k^I]$ initial code C^I and re-encoding the same data under a $[n^F, k^F]$ final code C^F. We call an instance of this problem a $[n^I, k^I; n^F, k^F]$ *convertible code*. Prior work on convertible codes have extensively studied their theoretical efficiencies under access cost and bandwidth cost, giving both lower bounds and optimal constructions.

However, these systems may be employed by nodes susceptible to malicious intruders, such as passive eavesdroppers or active adversaries, all of whom denigrate the conversion process. For instance, an eavesdropper may be able to learn the underlying data through the messages exchanged by nodes in the conversion process or an adversary may induce errors by sending malicious messages that cause the codeword to be corrupted. The setting of an insecure distributed storage system has been well studied for the related node repair problem [10]. **It is unknown whether optimal convertible codes for either access or bandwidth cost exist for insecure data storage systems.** In this paper, we bring the conversion problem into this setting, where we formalize the model of conversion in an insecure storage system, derive the maximum capacity of data that can be stored on such systems, and provide access and bandwidth optimal convertible code constructions that are

secure in such a system.

Chapter 2

Background and Relevant Work

In this chapter, we review relevant background and prior work. We will also define notation (emphasized in *italics*) on the way that we will use throughout the thesis.

2.1 Erasure Codes

Let $\mathbb{F}$ be a finite field of size q (q will be specified when relevant). An (n, k) erasure code $\mathcal{C}$ over $\mathbb{F}$ is a mapping from *messages* $\boldsymbol{m} \in \mathbb{F}^k$ to *codewords* $\boldsymbol{c} \in \mathbb{F}^n$. We say $\mathcal{C}$ is linear if it is a linear mapping and can be represented with *generator matrix* $\boldsymbol{G} \in \mathbb{F}^{k \times n}$ (we denote this by using square brackets e.g $\mathcal{C}$ is a $[n, k]$ code). We also say a $[n, k]$ erasure code $\mathcal{C}$ is systematic if $\boldsymbol{G} = \left[\boldsymbol{I}_k \mid \boldsymbol{P}\right]$, where $\boldsymbol{I}_k$ is the identity matrix of size k and we say $\boldsymbol{P}$ is the *parity matrix*.

Further, a $[n, k]$ code $\mathcal{C}$ is *Maximum Distance Separable (MDS)* if any subset of k columns are linearly independent (and thus form a non-singular matrix). In other words, any k symbols of a codeword is sufficient to reconstructing the entire codeword and its underlying message for systematic codes. In the next sections we will often refer to erasure codes simply as *codes*.

2.2 Convertible Codes

Convertible codes are a pair of codes designed for efficient encoding conversion[5, 6, 7, 8]. Specifically, let $\mathcal{C}^I$ be a (n^I, k^I) code over $\mathbb{F}$ and let $\mathcal{C}^F$ be a (n^F, k^F) code over $\mathbb{F}$. We consider the scenario where a data storage system has an initial, encoding the data under $\mathcal{C}^I$, and final configuration, encoding the data under code $\mathcal{C}^F$. The initial and final configurations may have differing code dimensions $k^I \neq k^F$, so the conversion problem considers multiple codewords in both the initial and final configuration. Specifically, we consider the storage of a message $\boldsymbol{m}$ with *length* $M = \text{lcm}(k^I, k^F)$, where there are $\lambda^I = M/k^I$ codewords from $\mathcal{C}^I$ in the initial configuration and $\lambda^F = M/k^F$ codewords from $\mathcal{C}^F$ in the final configuration. There must also be a mapping of data between initial and final configurations, specified by *partitions* $\mathcal{P}^I$ and $\mathcal{P}^F$ of $[M]$, each containing λ^I and λ^F subsets each respectively. Each $P_i^I \in \mathcal{P}^I$ has size $|P_i^I| = k^I$ and each $P_j^F \in \mathcal{P}^F$ has size $|P_j^F| = k^F$. We enumerate the codewords as follows: the submessage $\boldsymbol{m}|_{P_i^I}$, the projection of $\boldsymbol{m}$

to the symbols specified in P_i^I is encoded by initial codeword i, and the submessage $\boldsymbol{m}|_{P_j^F}$, the projection of $\boldsymbol{m}$ to the symbols specified in P_j^F is encoded by final codeword j. Then, conversion is defined as a procedure mapping initial codewords $\{\mathcal{C}^I(\boldsymbol{m}|_{P_i^F} : i \in [\lambda^I]\}$ to final codewords $\{\mathcal{C}^F(\boldsymbol{m}|_{P_j^F} : j \in [\lambda^F]\}$.

Definition 1 (Convertible Code)
A $(n^I, k^I; n^F, k^F)$ convertible code over $\mathbb{F}_q$ is defined by:

1. *A pair of codes $(\mathcal{C}^I, \mathcal{C}^F)$ where $\mathcal{C}^I$ is a (n^I, k^I) code over $\mathbb{F}$ and $\mathcal{C}^F$ is a (n^F, k^F) code over $\mathbb{F}_q$.*
2. *A pair of partitions $\mathcal{P}^I$ and $\mathcal{P}^F$ of $[M]$, where $M = lcm(k^I, k^F)$ such that each subset $P_i^I \in \mathcal{P}^I$ has size $|P_i^I| = k^I$, and each subset $P_j^F \in \mathcal{P}^F$ has size $|P_j^F| = k^F$.*
3. *A conversion procedure mapping $\{\mathcal{C}^I(\boldsymbol{m}|_{P_i^F}) : i \in [\lambda^I]\}$ to $\{\mathcal{C}^F(\boldsymbol{m}|_{P_j^F}) : j \in [\lambda^F]\}$.*

We say that a convertible code is MDS if the initial and final code are both MDS. Similarly, a convertible code is linear if the initial and final code are both linear.

Convertible codes have been largely explored within the contexts of access and bandwidth cost. The access cost of conversion is measured by the total number of nodes accessed in the process of conversion. Optimal bounds and constructions for the access cost of convertible codes is known for all valid $n^I, k^I, n^F, k^F \in \mathbb{N}^{>0}$ parameters ($n^I > k^I, n^F > k^F$).

Other work has also examined the bandwidth cost of conversion. Bandwidth cost of conversion is measured by the bandwidth used within a network of nodes employing conversion. Optimal bounds and constructions within the so-called merge regime ($k^F = \lambda^I k^F, \lambda^I \geq 2$) are known.

2.2.1 Access Optimal Convertible Code Constructions

Our secure convertible code constructions build off of existing access optimal convertible codes.

Example: Access Optimal $[7, 4; 8, 6]$ Convertible Code

Let θ be a primitive element in $\mathbb{F}$. An MDS, systematic, access optimal $[7, 4; 8, 6]$ convertible code is given by the following generators [1]

$$\mathcal{C}^I = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & \theta & \theta^2 \\ 0 & 0 & 1 & 0 & 1 & \theta^2 & \theta^4 \\ 0 & 0 & 0 & 1 & 1 & \theta^3 & \theta^6 \end{bmatrix}, \mathcal{C}^F = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & \theta \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & \theta^2 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & \theta^3 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & \theta^4 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & \theta^5 \end{bmatrix}$$

Note that the constructions are provided are generalize for any choice of k^I, k^F and $r^I \geq r^F$, the so-called *decreasing redundancy region*. In the case that $r^F > r^I$, the *increasing redundancy region*, the default conversion process is access optimal. If we allow a finer grained approach by allowing partial access of codeword symbols as sub-symbols, we can do better by making use of piggyback

[1] add the example footnote here.

codes[6, 7]. Additionally, one can observe that the field size requirement for such a construction is prohibitively high. Additional recent work has been done to find and characterize low field size constructions [1].

To our knowledge, all known constructions of optimal convertible codes are systematic. In the interest of 'securing' of the message, these codes are insufficient since they clearly reveal the message. We will devise codes that transform existing MDS convertible codes into ones that obfuscate the message.

2.3 Wiretap Channels

The *wiretap channel* was first introduced by Wyner in [12], where a transmitter Alice sends messages to receiver Bob across a discrete, memory-less channel in the presence of an Eavesdropper Eve. Wyner derives a tradeoff between the maximum rate of information that Alice can convey and the allowed amount of eavesdropping such that Alice's communication remains perfectly secret to Eve.

An extension of the wire-tap channel, known as the *wire-tap channel II*, was studied by Ozarow and Wyner in [9], where there is additional restrictions fixes Alice's message and encoding lengths. That is, Alice must convey a k symbol message to Bob over the channel with an n symbol transmission, where Eve can observe $\ell < n$ symbols. Ozarow and Wyner derive a similar upper bound on the information conveyed as in the previous wire-tap channel setting, but they also show an explicit construction matching the maximum information rate while remaining perfectly secret to any eavesdropper. Their construction is based on transmitting the cosets of a chosen *group code*, which will appear uniformly random to an eavesdropper who can only see a partial number of symbols. Since we adapt this technique for our setting of convertible codes, we delve into the technicalities in the following subsection 2.3.1.

Further, Subramanian and McLaughlin in [11] study the *erasure-erasure wire-tap channel*, a further extension of the wire-tap channel II with additional erasures μ in receiver Bob's view of Alice's sent transmission. Note that an erasure-erasure channel is a wire-tap II channel when $\mu = 0$. They construct a *nested code*, which is based on the coset encoding of Ozarow and Wyner, with an additional concatenation of an erasure code. Likewise, nested coding will be useful in constructing secure convertible codes, so we provide their background theory in the following subsection 2.3.1.

By including erasures in Bob's view, the erasure-erasure channel captures lead into secure lrc

2.3.1 Coset Binning and Nested Codes

Coset and nested coding used in the wiretap II channel[9] and the erasure-erasure channel[11] setting will be integral to our construction of secure convertible codes. We compile their relevant techniques in detail.

First, we look at coset codes in the wiretap II setting. Suppose Alice has a k symbol message $S \in \mathbb{F}^k$ she wants to transmit to Bob via an n symbol coded message $X \in \mathbb{F}^n$ through a perfect channel, where Eve may observe any $\ell < k$ symbols of her coded message. Alice's objective is to choose a n symbol encoding scheme that achieves perfect secrecy; that is, Eve does not gain any

information Alice's underlying message. We can state the requirements of the wiretap II setting in information theoretic terms as follows:

Definition 2 (Wiretap II Secure [9])
A code $\mathcal{C} \subseteq \mathbb{F}^n$ is (k, n, ℓ)-Wiretap II Secure if for any uniformly random chosen $S \in \mathbb{F}^k$ and $X \in \mathcal{C}$,

$$\begin{aligned} H(S|X_E) &= H(S), && (\forall E \subset [n], |E| \leq \ell) \\ H(S|X) &= 0. \end{aligned}$$

The first equation ensures that any ℓ symbols do not reveal anything about a message S chosen uniformly at random, and the second ensures that the entire message is recoverable. This is possible by employing the use of cosets as encodings of messages. Informally, a partial view of a coset vector admits candidate matches across different cosets, where each candidate coset contains an equal number of candidate vectors. Further, the number of candidate cosets will be equal to 2^k, the number of possible messages. Hence, an eavesdropper will have no information, while the receiver will be able to decode the entire message. Formally,

Lemma 3 (Coset Codes [9, 11])
For k, n, ℓ positive integers such that $k < n$ and $\ell \leq n - k$, there exists a code $\mathcal{C}_*$ that is (k, n, ℓ)-Wiretap II secure.

Proof. Choose $\mathcal{C}_*$ to be an MDS $[n, n-k]$ code. Since $|C_*| = q^{n-k}$, there are q^k cosets of $\mathcal{C}_*$. Suppose S is in some coset $s + \mathcal{C}_*$. Then, X is chosen to be a uniformly at random chosen element of the coset. Since there is a one-to-one correspondence of cosets to messages, X completely determines S, or $H(S|X) = 0$. What is left to show is that for any $E \subset [n], |E| \leq \ell, H(S|X_E) = H(S)$.

Let $a \in \mathbb{F}^n$ be a *match* for X_E. That is, for all $i \in E, a_i = (X_E)_i$. Then, $a + C_{[n]\setminus E}$ is the matches in a's corresponding coset, where

$$C_{[n]\setminus E} = \{c \in C_* : c_E = 0\}.$$

Then, the number of matches in each coset is $|C_{[n]\setminus E}|$, where there are $\frac{q^{n-|E|}}{|C_{[n]\setminus E}|}$ such cosets. Thus,

$$\begin{aligned} H(S|X_E) &= (n - |E|) - \dim C_{[n]\setminus E} \\ &= n - |E| - (n - k - |E|) = H(S). \end{aligned}$$

□

We move onto the erasure-erasure channel, where recall that this channel adds the addition of erasures in Bob's view. This addition is interesting due to push-and-pull between Bob's and Eve's goals. Now that Bob can only see some of the encoded message, enforcing recovery in his view while enforcing information theoretic security in Eve's view lowers the amount of information Alice can convey to Bob. As before, we will define (information theoretically) secure codes in the erasure-erasure channel, with the apparent addition of erasures in Bob's view and the less-apparent addition of Alice's reduced message length.

Definition 4 ((MDS) Erasure-Erasure Secure [11])
A code $\mathcal{C} \subseteq \mathbb{F}^n$ is (k, n, ℓ, μ)-Erasure-Erasure secure if for any uniformly random chosen $S \in \mathbb{F}^{k_S}$, and $X \in \mathcal{C}$,

$$\begin{aligned} H(S|X_E) &= H(S), && (\forall E \subset [n], |E| \leq \ell) \\ H(S|X_B) &= 0 && (\forall B \subset [n], |B| \geq \nu) \end{aligned}$$

for some $k_S \leq k$. If $\nu = k$, we say $\mathcal{C}$ is an MDS (k, n, ℓ)-Erasure-Erasure secure code.

There are two main changes: first, the modification of the second equation enforces that in the case of any $n - \nu$ erasures, the original message is still decodable. Second, each message is split into two parts $\begin{bmatrix} S & \kappa \end{bmatrix}$, where S is the k_S message symbols to be recovered by a decoder, and κ are k_κ redundant random symbols.

We denote the length of S to be the *secrecy capacity*. Using information theoretic arguments, the secrecy capacity can be upperbounded.

Theorem 5 (Upperbound on Secrecy Capacity for Erasure-Erasure Channels [11])
For k, n, ℓ, μ positive integers such that $\ell < \mu < k < n$, and code $\mathcal{C} : \mathbb{F}^k \to \mathbb{F}^n$, if $\mathcal{C}$ is a (k, n, ℓ, ν)-Erasure-Erasure secure code, then $k_s \leq \nu - \ell$.

Proof. Suppose $E \subset B \subset [n]$ such that $|E| = \ell$ and $|B| = \nu$. Then, for any uniformly random $S \in \mathbb{F}^{k_s}$, uniformly random $\kappa \in \mathbb{F}^{k_\kappa}$, and $X = C\left(\begin{bmatrix} S & \kappa \end{bmatrix}\right)$,

$$\begin{aligned} H(S) &= H(S|X_E) - H(S|X_B) \\ &= H(S|X_E) - H(S|X_E, X_{B\setminus E}) \\ &= I(S; X_{B\setminus E}|X_E) \\ &\leq H(X_{B\setminus E}|X_E) \\ &\leq H(X_{B\setminus E}) \\ &\leq \nu - \ell \end{aligned}$$

□

The intuition of the above bound is as follows: if Bob chooses some set of ν symbols to recover the message S, the worst case is when Eve chooses all her ℓ symbols from Bob's recovery set. In this case, the information conveyed by the ν symbols must be reduced by at least ℓ.

Are there Erasure-Erasure Secure codes that reach the maximum message length derived in theorem 5? While coset codes are not Erasure-Erasure secure for any $\ell > 0$ because cosets of MDS codes are not inherently recoverable from erasures, they are an important building block for a suitable construction known as the *nested code*.

Lemma 6 (Nested Codes [11])
For k, n, ℓ positive integers such that $\ell < k < n$, there exists a code D that is MDS (k, n, ℓ)-Erasure-Erasure secure with $k_s = k - \ell$.

Proof. Let $\mathcal{C}_S$ be a $[n, k - \ell]$ code and $\mathcal{C}_*$ be a MDS $[n, \ell]$ code such that $\mathcal{C}_S \cap C_* = \{0\}$ and $D = \mathcal{C}_S + \mathcal{C}_*$ is a MDS $[n, k]$ code. In other words, let G_S, G_*, and G be the generator matrices of

$\mathcal{C}_S$, MDS $\mathcal{C}_*$, and MDS D respectively. For any uniformly random $S \in \mathbb{F}^{k-\ell}$ and $\kappa \in \mathbb{F}^{\ell}$, we have

$$D\left(\begin{bmatrix} S & \kappa \end{bmatrix}\right) = \begin{bmatrix} S & \kappa \end{bmatrix} G = SG_S + \kappa G_*$$

Let $X = D\left(\begin{bmatrix} S & \kappa \end{bmatrix}\right)$. Since X is an element of some coset of $\mathcal{C}_*$, we can use the same analysis in lemma 3 to show that for any $J \subset [n]$ of revealed indices, we have

$$H(S|X_J) = \dim D_{[n]\setminus J} - \dim C_{*[n]\setminus J}{}^{2},$$

where for any $B \subset [n]$ of size $|B| = k$,

$$\dim D_{[n]\setminus B} - \dim C_{*[n]\setminus B} = (k-k) - 0.$$

Note that we use the fact that for any MDS $[n,k]$ code $\mathcal{C}$, $\dim \mathcal{C}_{[n]\setminus B} = \max\{0, k - |J|\}$. Lastly, for any $E \subset [n]$ of size $|E| = \ell$,

$$\dim D_{[n]\setminus E} - \dim C_{*[n]\setminus E} = (k-\ell) - (\ell - \ell) = k - \ell = H(S). \qquad \square$$

[2]There are a total of $|D_{[n]\setminus J}|$ matched cosets for X_J. Each matched coset will have $|\mathcal{C}_{*[n]\setminus J}|$ elements, where the number of cosets to consider are $\frac{|D_{[n]\setminus J}|}{|\mathcal{C}_{*[n]\setminus J}|}$.

Chapter 3

Eavesdropper-Secure Convertible Codes: Bound and Construction

In this chapter we define the eavesdropper model for convertible codes. In short, we augment the original framework for analyzing access cost of convertible codes with an additional Eavesdropper. The definition of a secure convertible code will be very general to start parameterized both by the number of eavesdropped symbols and the length of the true message stored securely. After deriving an upperbound on the secrecy capacity, we can drop the latter parameterization for a simpler definition.

3.1 Secure Convertible Codes: Modelling Conversion with an Eavesdropper

We introduce eavesdroppers in the convertible codes framework by using an erasure-erasure channel-inspired definition. We denote such convertible codes in which eavesdroppers learn information-theoretically nothing about the underlying message a *Secure Convertible Codes*.

We recall relevant convertible code notation and denote useful short-hands. Recall a convertible code is parameterized by positive integers k^I, n^I, k^F, n^F, where $\mathcal{M} = \text{lcm}\{k^I, k^F\}$ is the total number of symbols encoded, $\lambda^I = \mathcal{M}/k^I$ is the number of initial stripes, and $\lambda^F = \mathcal{M}/k^F$ is the number of final stripes. Further, for a convertible code $\mathcal{C} = (\mathcal{C}^I, \mathcal{C}^F)$ with partitions $(\mathcal{P}^I, \mathcal{P}^F)$, denote $\mathcal{C}^I(\boldsymbol{m}_{P^I}) = \left[\mathcal{C}^I\left(\boldsymbol{m}_{P^I_1}\right) \quad \dots \quad \mathcal{C}^I\left(\boldsymbol{m}_{P^I_{\lambda^I}}\right)\right]$ i.e the concatenation of each initial stripe of the convertible code. Define $\mathcal{C}^F(\boldsymbol{m}_{P^F})$ similarly.

Similar to prior work, we will be interested in convertible codes with maximum secrecy capacity, which we denote as $\mathcal{M}^S$. For the following definition, we will treat $\mathcal{M}^S$ as fixed, and we will derive the exact maximum secrecy capacity in theorem 9.

Definition 7 (ℓ-Secure Convertible Code)
$\mathcal{C} = (\mathcal{C}^I, \mathcal{C}^F)$ is an $\ell-$Secure $[n^I, k^I; n^F, k^F]$ Convertible Code if for any uniformly random $S \in \mathbb{F}^{\mathcal{M}_S}$, there exists encodings $X^I \in \mathbb{F}^{\lambda^I n^I}$, $X^F \in \mathbb{F}^{\lambda^F n^F}$, and partitions $\mathcal{P}^I = \{P^I_1, \dots, P^I_{\lambda^I}\}, \mathcal{P}^F = \{P^F_1, \dots, P^F_{\lambda^F}\}$ of $[\mathcal{M}^S]$ such that the following simultaneously hold:

1. ***Reconstruction.*** *For any $i \in [\lambda^I]$ and subset $B \subset P_i^I$ of size $|B| = k^I$,*

$$H(S_{P_i^I}|X_B^I) = 0,$$

and for any $j \in [\lambda^F]$ and subset $B \subset P_j^F$ of size $|B| = k^F$,

$$H(S_{P_j^F}|X_B^F) = 0.$$

2. ***ℓ-Secrecy.*** *For any $E^I \subset [\lambda^I n^I]$, $E^F \subset [\lambda^F n^F]$ of combined size $|E^I| + |E^F| \leq \ell$,*

$$H(S \mid X_{B_I}^I, X_{B_F}^F) = 0.$$

*Denote X^I as the **initial encoding** and X^F as the **final encoding** of S (under code $\mathcal{C}$ – this is omitted where the specified code is clear from context).*

To define eavesdroppers, our definition bridges the definition of traditional convertible codes to resemble erasure-erasure secure codes. First, similar to traditional convertible codes, we define initial and final configurations of the same message symbols under the implicitly defined encodings of X^I and X^F and partitions $\mathcal{P}^I$ and $\mathcal{P}^F$ such that reconstruction (MDS property) holds. With just the reconstruction property, our definition is equivalent to the original MDS convertible codes definition. Our definition augments the convertible codes framework with the second property. The ℓ-secrecy property dictates that any ℓ symbols between the initial and final stripe does not give any information about the underlying message S. Lastly, note that we consider the access cost of eavesdropper-secure convertible codes in the same manner as in traditional convertible codes. That is, the access cost is at least γ if there exists a conversion procedure mapping initial encoding X^I to final encoding X^F which accesses at most γ symbols of the initial encoding.

3.2 Secrecy Capacity in Secure Convertible Codes

We are interested in MDS Convertible codes that reach maximum secrecy capacity. Like the Erasure-Erasure channel, the property of (initial and final) reconstruction of convertible codes are at odds with eavesdropper security and thus bounds the secrecy capacity $\mathcal{M}_S$ stored by a convertible code. As a preliminary step towards deriving the maximum secrecy capacity, we, at a high level, observe that the secrecy property implies secrecy for each individual stripe.

Lemma 8 (Stripe-wise Secrecy of Secure Convertible Codes)
If $\mathcal{C}$ is an ℓ-secure $[n^I, k^I; n^F, k^F]$ convertible code with partitions $(\mathcal{P}^I, \mathcal{P}^F)$, then for uniformly random $S \in \mathbb{F}^{\mathcal{M}^S}$ with intial and final encodings $X^I \in \mathbb{F}^{\lambda^I n^I}, X^F \in \mathbb{F}^{\lambda^F n^F}$:

1. ***Initial Stripe Secrecy****: For $X^I = \mathcal{C}^I(\boldsymbol{m}_{\mathcal{P}^I})$ and for all $i \in [\lambda^I]$,*

$$H(S_{P_i^I}|X_E) = H(S_{\mathcal{P}_i^I}) \qquad (\forall E \subset I_i^I, |E| \leq \ell)$$

where $I_i^I = \{(i-1)n^I + 1, \ldots, in^I\}$.

2. ***Final Stripe Secrecy****: For all $j \in [\lambda^F]$,*

$$H(S_{P_j^F}|X_E) = H(S_{P_j^F}) \qquad (\forall E \subset I_j^F, |E| \leq \ell)$$

where $I_j^F = \{(j-1)n^F + 1, \ldots, jn^F\}$.

Proof. Without loss of generality, we consider initial secrecy. For $i \in [\lambda^I]$ and $E \subset I_i^I$ of size $|E| \leq \ell$, X_E only contains symbols in initial stripe i. Then if we assume for the sake of contradiction that $H(S_{P_i^I}|X_E) < H(S_{P_i^I})$, we have

$$\begin{aligned} H(S \mid X_E) &= \sum_{j=1}^{\lambda^I} H(S_{P_j^I}|X_E) \\ &= H(S_i|X_E) + \sum_{\substack{j=1 \\ j\neq i}}^{\lambda^I} H(S_{P_j^I}|X_E) \\ &\leq H(S_i|X_E) + \sum_{\substack{j=1 \\ j\neq i}}^{\lambda^I} H(S_{P_j^I}) \\ &< H(S_{P_i^I}) + \sum_{\substack{j=1 \\ j\neq i}}^{\lambda^I} H(S_{P_j^I}) = H(S), \end{aligned}$$

and we reach a contradiction by the ℓ-secrecy property of $\mathcal{C}$ is an ℓ-secure convertible code. □

We are now prepared to derive the upperbound on the secrecy capacity for MDS convertible codes. We follow the proof used in 5 along with lemma 8.

Theorem 9 (Secrecy Capacity of MDS Convertible Codes)
For positive integers k^I, n^I, k^F, n^F, ℓ such that $k^I \leq n^I, k^F \leq n^F$, $\ell < \min\{k^I, k^F\}$, if $\mathcal{C}$ is an $\ell-$secure $[n^I, k^I; n^F, k^F]$ convertible code, then

$$\mathcal{M}_S \leq \min\{\lambda^I(k^I - \ell), \lambda^F(k^F - \ell)\}.$$

Proof. Suppose $k^I \leq k^F$. Fix $i \in [\lambda^I]$ and suppose $E \subset B \subset I_i^I$ such that $|E| = \ell$ and $|B| = k^I$. Then,

$$\begin{aligned} H(S_{P_i^I}) &= H(S_{P_i^I}|X_E) - H(S_{P_i^I}|X_B) && \text{(lem. 8)} \\ &= H(S_{P_i^I}|X_E) - H(S_{P_i^I}|X_E, X_{B\setminus E}\} \\ &= I(S_{P_i^I}; X_{B\setminus E}|X_{E_i}) \\ &\leq H(X_{B\setminus E}|X_E) \\ &\leq H(X_{B\setminus E}) \\ &\leq k^I - \ell, \end{aligned}$$

where

$$\mathcal{M}_S = H(S) = \sum_{i=1}^{\lambda^I} H(S_{P_i^I}) \leq \lambda^I(k^I - \ell).$$

Now, suppose $k^F < k^I$. Fix $j \in [\lambda^F]$ and suppose $E \subset B \subset I_F^I$ such that $|E| = \ell$ and $|B| = k^F$. Then, symmetric to the previous case, we have $H(S_{P_j^F}) \leq k^F - \ell$ and $\mathcal{M}_S \leq \lambda^F(k^F - \ell)$. Putting the two cases together, we have our desired bound. □

The intuition for the secrecy capacity of convertible codes is that our model essential puts $\lambda^I + \lambda^F$ encodings of the same partitioned message, the initial and final stripes, through an Erasure-Erasure channel. Since these encodings are a concatenation of MDS code stripes, the decoder must chooses λ^I (resp. λ^F) subsets of each initial (resp. final) stripe. The best an eavesdropper can do is to choose to eavesdrop all their ℓ symbols in a particular stripe's subset. This is because the symbols encoded in one stripe does not give any information about the symbols encoded in another stripe. Thus, a secure convertible code is only possible if it handles ℓ-eavesdropped symbols on each stripe.

Note that an interesting extension of the above intuition is to consider if knowing the number of eavesdropper per stripe would improve the secrecy capacity. We denote the maximum information stored in this setting as *fine-grained secrecy capacity*. Likewise, codes which satisfy this setting are denoted as *fine-grained secure convertible codes*. Fine-grained secure conversion is explored further in chapter 4.

In the next chapter, we show that it is possible to construct secure convertible codes for all valid parameters by bootstrapping existing access-optimal convertible codes with nested codes.

In this chapter we construct access optimal convertible codes with optimal secrecy capacity.

3.3 A Construction for General Parameters

Our technique will be to use existing access optimal convertible codes along with a nested code. To provide intuition on the coding scheme, we are essentially masking the message with random symbols, which we then use as the message for the convertible code. After conversion, the final encoding will be over the masked message, which we will have the additional step of unmasking. We will be able to ensure this unmasking by showing that the final encoding will encode each partition with a nested code.

Theorem 10 (Explicit Construction of Convertible Codes with Optimal Secrecy Capacity)
For any parameters n^I, n^F, k^I, k^F, there exists an access-cost optimal ℓ-secure $[n^I, k^I; n^F, k^F]$ convertible code with maximum secrecy capacity $\mathcal{M}_S$ for $\ell < \min\{k^I, k^F\}$.

Proof. By the previous theorem 9, the secrecy capacity is upperbounded by $\mathcal{M}_S \leq \min\{\lambda^I(k^I - \ell), \lambda^F(k^F - \ell)\}$. Without loss of generality, suppose that $\lambda^I(k^I - \ell) \geq \lambda^F(k^F - \ell)$ so $\mathcal{M}_S = \lambda^I(k^I - \ell)$. [1]

First, let $S \in \mathbb{F}^{\mathcal{M}_S}$ be the message to be secured and let $\kappa \in \mathbb{F}^{\lambda^I \ell}$ be random symbols. By [8], there exists an access optimal convertible code $(\mathcal{C}^I, \mathcal{C}^F)$ with partitions $(\mathcal{P}^I, \mathcal{P}^F)$ with the desired parameters. Based on these partitions, form the i'th initial message $\boldsymbol{m}_i^I$ for each $i \in [\lambda^I]$ as

$$\boldsymbol{m}_i^I = \begin{bmatrix} \kappa & S[(i-1)(k^I - \ell) + 1 : i(k^I - \ell)] \end{bmatrix}.$$

There is a slight modification for final messages. We form the j'th final message $\boldsymbol{m}_j^F$ for $j < \lambda^F$, that is each final message except the last, as

$$\boldsymbol{m}_j^F = \begin{bmatrix} \kappa & S[(j-1)(k^F - \ell) + 1 : j(k^F - \ell)] \end{bmatrix}.$$

[1] Note that this is equivalent to when $\lambda^I \geq \lambda^F$ i.e the 'generalized merge regime.' In the case $\lambda^F \geq \lambda^I$ (the 'generalized split regime.'), the construction will be 'reversed' in a straight-forward manner.

For the last final message, choose $\lambda^F(k^F - k^I - \ell)$ symbols arbitrarily from used symbols in S. Call these symbols S'. The last message is formed as

$$\boldsymbol{m}^F_{\lambda^F} = \begin{bmatrix} \kappa & S[(\lambda^F - 1)(k^F - \ell) + 1 : \lambda^F(k^F - \ell)] & S' \end{bmatrix}$$

Now, we are prepared to describe our convertible code with maximum secrecy capacity. Consider a MDS nested $[k^I, k^I]$ code $D^I = \begin{bmatrix} D^I_\kappa \\ D^I_{\boldsymbol{m}} \end{bmatrix}$ where D^I_κ is a MDS $[k^I, \ell]$ code. Symmetrically, define D^F except with final parameters. Then, we form our initial code as a concatenated code $\mathcal{C}^I \circ D^I$ and our final code as $\mathcal{C}^F \circ D^F$. Since the concatenation of MDS codes are MDS, the constructed codes are MDS. Lastly, our codes are perfectly secret using the same reasoning in the erasure-erasure channel case[11]. □

3.3.1 Example: 1-secure $[5, 4; 7, 6]$ Convertible Code

For example, we illustrate an MDS $1-$private $[5, 4; 7, 6]$ convertible code ($\ell = 1$). Here, $\lambda^I = 3$ and $\lambda^F = 2$, so the maximum secrecy capacity is $\min\{3(4-1), 2(6-1)\} = 9$. Let $m = m_1 \dots m_9$ be the message symbols and let $\kappa, \kappa' \in \mathbb{F}$ be a uniformly random chosen symbol.

Then, our initial messages are

$$\begin{bmatrix} \kappa & m_1 & m_2 & m_3 \end{bmatrix},$$
$$\begin{bmatrix} \kappa & m_4 & m_5 & m_6 \end{bmatrix},$$
$$\begin{bmatrix} \kappa' & m_7 & m_8 & m_9 \end{bmatrix},$$

with final messages,

$$\begin{bmatrix} \kappa & m_1 & m_2 & m_3 & m_4 & m_5 \end{bmatrix},$$
$$\begin{bmatrix} \kappa & \kappa & m_6 & m_7 & m_8 & m_9 \end{bmatrix}.$$

We choose the MDS $[4, 4]$ nested code $\mathcal{D}^I$ with generator

$$\mathcal{D}^I = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{2}.$$

Take any optimal MDS $[7, 6; 5, 4]$ convertible code $(\mathcal{C}^I, \mathcal{C}^F)$ with partitions specified by the process below. Then, our initial code will have generator $\mathcal{D}^I \times \mathcal{C}^I$, that is we have a concatenated code with inner (first) code $\mathcal{D}^I$ and outer code $\mathcal{C}^I$. Then, the generated initial codewords are

$$\begin{bmatrix} \kappa & m_1 & m_2 & m_3 \end{bmatrix} \left(\mathcal{D}^I \times \mathcal{C}^I\right) = \begin{bmatrix} \kappa & \hat{m}_1 & \hat{m}_2 & \hat{m}_3 \end{bmatrix} \mathcal{C}^I,$$
$$\begin{bmatrix} \kappa & m_4 & m_5 & m_6 \end{bmatrix} \left(\mathcal{D}^I \times \mathcal{C}^I\right) = \begin{bmatrix} \kappa & \hat{m}_4 & \hat{m}_5 & \hat{m}_6 \end{bmatrix} \mathcal{C}^I,$$
$$\begin{bmatrix} \kappa & m_7 & m_8 & m_9 \end{bmatrix} \left(\mathcal{D}^I \times \mathcal{C}^I\right) = \begin{bmatrix} \kappa & \hat{m}_7 & \hat{m}_8 & \hat{m}_9 \end{bmatrix} \mathcal{C}^I,$$

[2]For the sake of notational simplicity, we will overload notation for codes with their generators.

where $\hat{m}_i = m_i + \kappa$. We use as-is the conversion method of the underlying convertible code. $(\mathcal{C}^I, \mathcal{C}^F)$, and after conversion, we have final codewords

$$\begin{bmatrix}\kappa & \hat{m}_1 & \hat{m}_2 & \hat{m}_3 & \hat{m}_4 & \hat{m}_5\end{bmatrix}\mathcal{C}^F,$$
$$\begin{bmatrix}\kappa & \kappa & \hat{m}_6 & \hat{m}_7 & \hat{m}_8 & \hat{m}_9\end{bmatrix}\mathcal{C}^F.$$

These final codewords are secure to any eavesdropped symbol, since

$$\begin{bmatrix}\kappa & \hat{m}_1 & \hat{m}_2 & \hat{m}_3 & \hat{m}_4 & \hat{m}_5\end{bmatrix}\mathcal{C}^F = \begin{bmatrix}\kappa & m_1 & m_2 & m_3 & m_4 & m_5\end{bmatrix}(\mathcal{D}_1^F \times \mathcal{C}^F),$$
$$\begin{bmatrix}\kappa & \kappa & \hat{m}_6 & \hat{m}_7 & \hat{m}_8 & \hat{m}_9\end{bmatrix}\mathcal{C}^F = \begin{bmatrix}\kappa & \kappa & m_6 & m_7 & m_8 & m_9\end{bmatrix}(\mathcal{D}_2^F \times \mathcal{C}^F).$$

where $\mathcal{D}_1^F, \mathcal{D}_2^F$ are nested $[6, 6]$ codes with generators

$$\mathcal{D}_1^F = \begin{bmatrix}1&1&1&1&1&1\\0&1&0&0&0&0\\0&0&1&0&0&0\\0&0&0&1&0&0\\0&0&0&0&1&0\\0&0&0&0&0&1\end{bmatrix}, \mathcal{D}_2^F = \begin{bmatrix}1&0&1&1&0&0\\0&1&0&0&1&1\\0&0&1&0&0&0\\0&0&0&1&0&0\\0&0&0&0&1&0\\0&0&0&0&0&1\end{bmatrix}.$$

It remains to show how we decode our final codewords from any k^F symbols. By the MDS property of $\mathcal{C}^F$, reading any k^F symbols recovers the nested messages,

$$\begin{bmatrix}\kappa & \hat{m}_1 & \hat{m}_2 & \hat{m}_3 & \hat{m}_4 & \hat{m}_5\end{bmatrix},$$
$$\begin{bmatrix}\kappa & \kappa & \hat{m}_6 & \hat{m}_7 & \hat{m}_8 & \hat{m}_9\end{bmatrix}.$$

Thus, we can recover the desired final message symbols by subtracting κ from each $\hat{m}_i$ symbol.

Remark. Note that in traditional convertible codes, we require that the initial and final codes to be the same for every stripe. This example demonstrates that the nested coding may be different per stripe. Nonetheless, as mentioned earlier, if we view the secure encoding procedure to be a preprocessing/postprocessing for intial/final configurations respectively, the convertible code used can still be fixed as a single initial and a single final code.

3.3.2 Example: 2-secure $[5, 3; 7, 6]$ Convertible Code

Chapter 4

Fine-grained Secure Convertible Codes

4.1 Increasing Capacity with Known Eavesdropper distribution

A significant constraint on the information capacity is the lack of information on where Eve's ℓ eavesdropped symbols lie. Concretely, consider a merge convertible code with merge parameter λ^I. In the worst case, all of Eve's ℓ eavesdropped symbols may lie in a single codeword. Intuitively, this is reflected in the capacity upperbound, which treats every initial stripe as if each contained ℓ eavesdropped symbols. A natural question is if we knew exactly how many symbols Eve has eavesdropped on each individual stripe, could we do better? In this section, we answer with an affirmative, providing both a lower bound and matching construction. Note that this assumption captures an interesting scenario where the number of eavesdropped symbols in each stripe either increase or decrease. Figure 4.1 illustrates an example of this in the natural setting of "merging eavesdroppers" and "splitting eavesdroppers." We proceed with modifying our existing eavesdropped convertible code definition to state how many eavesdropped symbols exist per intial and final stripe.

Definition 11 (($\{\ell_i^I\}_{i\in[\lambda^I]}, \{\ell_j^F\}_{j\in[\lambda^F]}$)$-$Secure $[n^I, k^I; n^F, k^F]$ Convertible Code)
$\mathcal{C}$ is an $(\{\ell_i^I\}_{i\in[\lambda^I]}, \{\ell_j^F\}_{j\in[\lambda^F]})-$Secure $[n^I, k^I; n^F, k^F]$ Convertible Code if,

1. *$\mathcal{C}$ is a $[n^I, k^I; n^F, k^F]$ convertible code,*
2. *Let $S \in \mathbb{F}^{\mathcal{M}_S}$ be uniformly random. Let $X^I \in \left(\mathcal{C}^I\right)^{\lambda^I}$, and $X^F = \left(\mathcal{C}^F\right)^{\lambda^F}$. For any chosen index sets $\{\mathcal{I}_i^I\}_{i\in[\lambda^I]}$, and $\{\mathcal{I}_j^F\}_{j\in[\lambda^F]}$*

$$H(S|X^I_{\mathcal{I}^I_1}, \ldots, X^I_{\mathcal{I}^I_{\lambda^I}}, X^F_{\mathcal{I}^F_1}, \ldots, X^F_{\mathcal{I}^F_{\lambda^F}}) = H(S),$$

   such that for all $i \in [\lambda^I].j \in [\lambda^F]$, we have $\mathcal{I}_i^I \subset [(i-1)n^I+1, in^I], \mathcal{I}_j^F \subset [(j-1)n^F+1, in^F]$ and $|\mathcal{I}_i^I| \leq \ell_i^I, |\mathcal{I}_j^F| \leq \ell_j^F$.
3. *$\sum_{i=1}^{\lambda^I} \ell_i^I + \sum_{j=1}^{\lambda^F} \ell_j^F = \ell$.*

Theorem 12 (Fine-Grained Secrecy Capacity of Convertible Codes)
For positive integers k^I, n^I, k^F, n^F, and $\ell_1^I, \ldots, \ell_{\lambda^I}^I, \ell_1^F, \ldots, \ell_{\lambda^F}^F$ such that $k^I \leq n^I, k^F \leq n^F$, $0 \leq \ell_1^I, \ldots, \ell_{\lambda^I}^I < k^I, 0 \leq \ell_1^F, \ldots, \ell_{\lambda^F}^F < k^F$, and convertible code $\mathcal{C}$, if $\mathcal{C}$ is an $(\{\ell_i^I\}_{i\in[\lambda^I]}, \{\ell_j^F\}_{j\in[\lambda^F]})-$secure

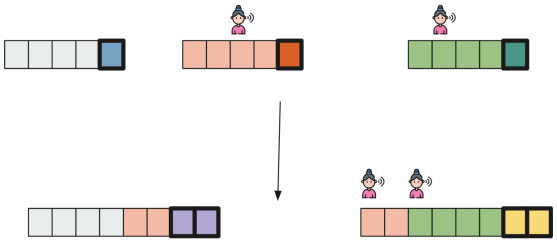

Figure 4.1: A $(\{0,1,1\},\{0,2\})$-Secure $[5,4;8,6]$ Convertible Code can model the scenario where long-term eavesdroppers may be 'merged' into the same stripe. If we were to use a $4-$Secure $[5,4;8,6]$ Convertible Code, then the secrecy capacity would be 0 and there is no secure data stored on a system. However, a $(\{0,1,1\},\{0,2\})$-Secure $[5,4;8,6]$ Convertible Code may store 10 symbols.

$[n^I,k^I;n^F,k^F]$ *convertible code, then*

$$\mathcal{M}_S \leq \min\left\{\sum_{i=1}^{\lambda^I}(k^I-\ell_i^I),\sum_{j=1}^{\lambda^F}(k^F-\ell_j^F)\right\}.$$

Proof. Suppose $k^I \leq k^F$. For each $i \in [\lambda^I]$, suppose $E_i \subset B_i \subset I_i^I$ such that $|E_i| = \ell_i^I$ and $|B_i| = k^I$.

Then,

$$\begin{aligned} H(S_{P_i^I}) &= H(S_{P_i^I}|X_{E_i}) - H(S_{P_i^I}|X_{B_i}) && \text{(cor. ??, lem. 8)}\\ &= H(S_{P_i^I}|X_{E_i}) - H(S_{P_i^I}|X_{E_i},X_{B_i\setminus E_i}\}\\ &= I(S_{P_i^I};X_{B_i\setminus E_i}|X_{E_i})\\ &\leq H(X_{B_i\setminus E_i}|X_{E_i})\\ &\leq H(X_{B_i\setminus E_i})\\ &\leq k^I-\ell_i^I, \end{aligned}$$

where

$$\mathcal{M}_S = H(S) = \sum_{i=1}^{\lambda^I} H(S_{P_i^I}) \leq \sum_{i=1}^{\lambda^I}(k^I-\ell_i^I).$$

The modification for $k^F < k^I$ is symmetrical. □

Note that if we take $\ell = \max\left\{\sum_{i\in[\lambda^I]}\ell_i^I, \sum_{j\in[\lambda^F]}\ell_j^F\right\}$ then we equivalently have that for fine-grained convertible codes, $\mathcal{M}_S \leq \mathcal{M}-\ell$. That is, with fine-grained security, we remove the restrictive λ^I or λ^F term in the original secrecy capacity.

4.1.1 Merging Eavesdroppers Example

4.1.2 Example: 1-merge-secure $[4, 2; 6, 4]$ Convertible Code

There are $2(2-1) = 2$ message symbols and $\lambda\ell = 2$ random symbols. The messages are

$$\begin{bmatrix} \kappa_1 & m_1 \end{bmatrix}$$
$$\begin{bmatrix} \kappa_2 & m_2 \end{bmatrix}$$

We use an initial code $\mathcal{C}^I$ of a $1-$secure convertible code as our initial encoding, a concatenation of a $[2, 2]$ nested code and an access optimal $[4, 2]$ initial code. For simplicity, we use the original access optimal merge regime code with primitive element θ to generate the parities.

$$\mathcal{C}^I\left(\begin{bmatrix} \kappa_1 & m_1 \end{bmatrix}\right) = \begin{bmatrix} \kappa_1 & m_1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & \theta \end{bmatrix} = \begin{bmatrix} \kappa_1 + m_1 & \kappa_1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & \theta \end{bmatrix}$$
$$\mathcal{C}^I\left(\begin{bmatrix} \kappa_2 & m_2 \end{bmatrix}\right) = \begin{bmatrix} \kappa_2 & m_2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & \theta \end{bmatrix} = \begin{bmatrix} \kappa_2 + m_2 & \kappa_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & \theta \end{bmatrix}$$

Then, we read the $\lambda^I r^F = 2$ parities and the $\lambda^I \ell = 2$ random symbols κ_1, κ_2. If we follow the original $1-$secure scheme we would have final codeword

$$\begin{bmatrix} \kappa_1 & \kappa_2 & \kappa_1 + m_1 & \kappa_2 + m_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & \theta \\ 0 & 0 & 1 & 0 & 1 & \theta^2 \\ 0 & 0 & 0 & 1 & 1 & \theta^3 \end{bmatrix}$$

However, this is insufficient for 2 eavesdroppers. Instead, the idea is that we generate final codeword

$$\begin{bmatrix} \kappa_1 + \boldsymbol{\kappa_2} & \boldsymbol{2\kappa_1} + \kappa_2 & m_1 + \kappa_1 & m_2 + \kappa_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & \theta \\ 0 & 0 & 1 & 0 & 1 & \theta^2 \\ 0 & 0 & 0 & 1 & 1 & \theta^3 \end{bmatrix}.$$

See that this is secure against 2 eavesdroppers because

$$\begin{bmatrix} \kappa_1 + \boldsymbol{\kappa_2} & \boldsymbol{2\kappa_1} + \kappa_2 & m_1 + \kappa_1 & m_2 + \kappa_2 \end{bmatrix} = \begin{bmatrix} \kappa_1 & \kappa_2 & \kappa_1 + m_1 & \kappa_2 + m_2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and we can observe the matrix forms a nested $[4, 4]$ code. Thus, the final codeword will be

$$\begin{bmatrix} \kappa_1 & \kappa_2 & \kappa_1 + m_1 & \kappa_2 + m_2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & \theta \\ 0 & 0 & 1 & 0 & 1 & \theta^2 \\ 0 & 0 & 0 & 1 & 1 & \theta^3 \end{bmatrix}$$

What is left to show is that we can generate the parities. Using these symbols, we can generate the first parity as,

$$\begin{aligned}\begin{bmatrix}\kappa_1 & \kappa_1+m_1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}+\begin{bmatrix}\kappa_2 & \kappa_2+m_2\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}+\boldsymbol{2\kappa_1+\kappa_2}&\\ =\kappa_1+(\kappa_1+m_1)+\kappa_2+(\kappa_2+m_2)+\boldsymbol{2\kappa_1+\kappa_2}&\\ =(\kappa_1+\boldsymbol{\kappa_2})+(\boldsymbol{2\kappa_1}+\kappa_2)+(m_1+\kappa_1)+(m_2+\kappa_2)&\\ =\begin{bmatrix}\kappa_1+\boldsymbol{\kappa_2} & \boldsymbol{2\kappa_1}+\kappa_2 & m_1+\kappa_1 & m_2+\kappa_2\end{bmatrix}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}&\end{aligned}$$

and the second parity as

$$\begin{aligned}\begin{bmatrix}\kappa_1 & \kappa_1+m_1\end{bmatrix}\begin{bmatrix}1\\\theta\end{bmatrix}+\theta^2\begin{bmatrix}\kappa_2 & \kappa_2+m_2\end{bmatrix}\begin{bmatrix}1\\\theta\end{bmatrix}+\boldsymbol{2\kappa_1\theta^2+\kappa_2}&\\ =(\kappa_1+\boldsymbol{\kappa_2})+\theta(\kappa_1+m_1)+\theta^2(\boldsymbol{2\kappa_1}+\kappa_2)+\theta_3(\kappa_2+m_2)&\\ =\begin{bmatrix}\kappa_1+\boldsymbol{\kappa_2} & \boldsymbol{2\kappa_1}+\kappa_2 & m_1+\kappa_1 & m_2+\kappa_2\end{bmatrix}\begin{bmatrix}1\\\theta\\\theta^2\\\theta^3\end{bmatrix}&\end{aligned}$$

4.1.3 Splitting Eavesdroppers Example

4.1.4 Increasing Vulnerability Example

The following is a $2-$secure $[5,3;7,6]$ convertible code where we know that $\ell_1=\ell_2=1$. If we allow ourselves two symbols of randomness, $\kappa=\kappa_1\kappa_2$, then we have initial messages

$$\begin{bmatrix}\kappa_1 & m_1 & m_2\end{bmatrix}$$
$$\begin{bmatrix}\kappa_2 & m_3 & m_4\end{bmatrix}$$

and final message

$$\begin{bmatrix}\kappa_1 & \kappa_2 & m_1 & m_2 & m_3 & m_4\end{bmatrix}$$

Then, we can use a $1-$secure convertible code.

Chapter 5

Conclusion

Bibliography

[1] Saransh Chopra, Francisco Maturana, and K. V. Rashmi. On low field size constructions of access-optimal convertible codes, 2024. URL https://arxiv.org/abs/2405.09010. 2.2.1

[2] Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung. The google file system. *SIGOPS Oper. Syst. Rev.*, 37(5):29â43, oct 2003. ISSN 0163-5980. doi: 10.1145/1165389.945450. URL https://doi.org/10.1145/1165389.945450. 1

[3] Cheng Huang, Huseyin Simitci, Yikang Xu, Aaron Ogus, Brad Calder, Parikshit Gopalan, Jin Li, and Sergey Yekhanin. Erasure coding in windows azure storage. In *2012 USENIX Annual Technical Conference (USENIX ATC 12)*, pages 15–26, Boston, MA, June 2012. USENIX Association. ISBN 978-931971-93-5. URL https://www.usenix.org/conference/atc12/technical-sessions/presentation/huang. 1

[4] Saurabh Kadekodi, K. V. Rashmi, and Gregory R. Ganger. Cluster storage systems gotta have heart: improving storage efficiency by exploiting disk-reliability heterogeneity. In *Proceedings of the 17th USENIX Conference on File and Storage Technologies*, FAST'19, page 345â358, USA, 2019. USENIX Association. ISBN 9781931971485. 1

[5] Francisco Maturana and K. V. Rashmi. Convertible codes: Efficient conversion of coded data in distributed storage, 2019. 1, 2.2

[6] Francisco Maturana and K. V. Rashmi. Bandwidth cost of code conversions in distributed storage: Fundamental limits and optimal constructions, 2020. 1, 2.2, 2.2.1

[7] Francisco Maturana and K. V. Rashmi. Bandwidth cost of code conversions in the split regime, 2022. 1, 2.2, 2.2.1

[8] Francisco Maturana, V. S. Chaitanya Mukka, and K. V. Rashmi. Access-optimal linear mds convertible codes for all parameters, 2020. 1, 2.2, 3.3

[9] L. H. Ozarow and A. D. Wyner. Wire-tap channel ii. *ATT Bell Laboratories Technical Journal*, 63(10):2135–2157, 1984. doi: 10.1002/j.1538-7305.1984.tb00072.x. 2.3, 2.3.1, 2, 3

[10] Sameer Pawar, Salim El Rouayheb, and Kannan Ramchandran. Securing dynamic distributed storage systems against eavesdropping and adversarial attacks, 2011. 1

[11] Arunkumar Subramanian and Steven W. McLaughlin. Mds codes on the erasure-erasure wiretap channel, 2009. 2.3, 2.3.1, 3, 4, 5, 6, 3.3

[12] A. D. Wyner. The wire-tap channel. *The Bell System Technical Journal*, 54(8):1355–1387,

1975. doi: 10.1002/j.1538-7305.1975.tb02040.x. 2.3