Nearly Optimal Robust Secret Sharing

Mahdi Cheraghchi*
Department of Computing
Imperial College London
London, UK

Abstract

We prove that a known approach to improve Shamir's celebrated secret sharing scheme; i.e., adding an information-theoretic authentication tag to the secret, can make it robust for n parties against any collusion of size δn, for any constant $\delta \in (0, 1/2)$. This result holds in the so-called "non-rushing" model in which the n shares are submitted simultaneously for reconstruction. We thus obtain an efficient and robust secret sharing scheme in this model that is essentially optimal in all parameters including the share size which is $k(1 + o(1)) + O(\kappa)$, where k is the secret length and κ is the security parameter. Like Shamir's scheme, in this modified scheme any set of more than δn honest parties can efficiently recover the secret.

Using algebraic geometry codes instead of Reed-Solomon codes, we decrease the share length to a constant (only depending on δ) while the number of shares n can grow independently. In this case, when n is large enough, the scheme satisfies the "threshold" requirement in an approximate sense; i.e., any set of $\delta n(1 + \rho)$ honest parties, for arbitrarily small $\rho > 0$, can efficiently reconstruct the secret.

1 Introduction

Secret sharing, introduced by the seminal works of Shamir [Sha79] and Blackley [Bla79], is the following problem (in its most basic formulation): Suppose we wish to encode and distribute a secret $s \in \mathbb{F}_2^k$ among n parties in such a way that i) the n parties can reconstruct the original secret s by revealing their respective shares; and, ii) for some integer parameter $t > 0$ (called the *privacy parameter*), any group of t parties cannot infer any information about the secret from their collection of shares. In coding-theoretic terms, the goal is to encode s (using randomness) into a sequence $Y_1, \ldots, Y_n$ over some alphabet of size Q, in a way that s can be reconstructed from the encoding and moreover, for any $i_1, \ldots, i_t \in [n]$, the sequence $Y_{i_1}, \ldots, Y_{i_t}$ has the same distribution regardless of the message s.

*Email: ⟨m.cheraghchi@imperial.ac.uk⟩.

Shamir proposed a beautiful scheme that provides an optimal solution to the problem. The scheme regards the secret as an element of the finite field $\mathbb{F}_Q$, for some prime power $Q \geq n$, and then samples a uniformly random univariate polynomial of degree t over $\mathbb{F}_Q$ with the constant term set to be s. The coding-theoretic interpretation of this solution is that s is amended with t uniformly random and independent elements of $\mathbb{F}_Q$ and the result is encoded using a Reed-Solomon code of length n and dimension $t+1$. Shamir's solution works even if the adversary uses an adaptive strategy; i.e., when each of the query positions $i_1, \ldots, i_t$ depends on the observation outcomes at the previous locations. Adaptive security is a property that is generally sought after for secret sharing schemes.

Due to its coding-theoretic nature, Shamir's scheme provides at least two additional benefits. First, any group of parties is able to recover s as long as the size of the group is larger than t. This so-called "threshold property" is due to the fact that the Reed-Solomon code is an MDS code. Second, any Reed-Solomon code of rate R is able to tolerate any fraction of errors up to $(1-R)/2$ and this can be achieved by an efficient decoder (such as the Berlekamp–Massey decoding algorithm, cf. [Rot06, Chapter 6]). As a result, a straightforward calculation shows that Shamir's secret sharing scheme is *robust*, in the sense that it can tolerate any less than $1/3$ fraction of *dishonest* parties. That is, the correct secret s can be reconstructed even if any less than $1/3$ fraction of the parties reveal their shares incorrectly. In fact, this holds true even if the malicious parties are able to arbitrarily communicate with each other and choose the incorrect shares adversarially.

More strongly, Shamir's scheme is secure against the so-called "rushing" adversaries. In the rushing setting (also known as "secret sharing with reconstructor"), reconstruction is done by each party broadcasting their (possibly corrupted) shares in an order determined by the protocol. This means that the adversary may attempt to, adaptively, manipulate shares at any point in the reconstruction phase (up to its allotted budget) based on its (adaptive) observation of up to t shares as well as all the shares (including those of the honest parties) that are revealed so far. Naturally, the requirement is then that each party should be able to correctly reconstruct the secret in isolation, with high probability, from the information received from the n parties. The error resilience of Shamir's scheme is based on the minimum distance of Reed-Solomon code, and thus the power of the adversary is irrelevant for this scheme as long as the number of manipulations is less than the minimum distance of the code. In fact in the reconstruction phase the adversary may observe *everyone's* shares and then decide which ones to corrupt, and the set of corrupted shares may or may not overlap with the set of t shares observed by the adversary before reconstruction (an interesting property that is not in general required in robust secret sharing, but is nevertheless satisfied by some known constructions that rely on error-correcting codes to provide robustness; e.g., [SNW15]).

1.1 Previous work

The robust notion of secret sharing has been studied in the literature, and some of the key results in the area are summarized in Table 1. It is known that robust secret sharing is impossible when the fraction of dishonest parties is at least $1/2$ [IOS12]. It is also impossible to always reconstruct the secret correctly (i.e., with probability 1) when the fraction of dishonest parties may be $1/3$ or larger, in which case a small probability of error η is unavoidable. Therefore, Shamir's scheme provides optimal robustness for a scheme with zero probability of error.

When an honest majority exists, Rabin and Ben-Or [RBO89] provide a secret sharing scheme based on Shamir's scheme combined with message authentication codes. The share length $q := \log Q$ in this scheme is, ignoring small terms, $k+\Omega(n\log(1/\eta))$, where $\eta > 0$ is the probability of incorrect reconstruction. In contrast, an appealing feature of Shamir's scheme is that the shares are *compact*; namely, the bit length of each share is equal to the bit length of the secret (under the natural assumption that $n \leq 2^k$). This turns out to be optimal for schemes with perfect privacy satisfying the threshold property [Sti92].

Another scheme, due to Cramer et al. [CDF01] (and based on [CPS99] and also using Shamir's scheme) improves the share length to $\max\{k, O(n+\log(1/\eta))\}$. However, the reconstruction time for this scheme is in general exponential in n (more precisely, at least $\binom{n}{t}$), and the scheme is insecure against rushing adversaries (cf. [CFOR12]).

Cevallos et al. [CFOR12] propose a scheme similar to [RBO89] that achieves more compact shares, namely of length $k + O(\log(1/\eta) + n(\log n + \log k))$. This scheme provides efficient share and reconstruction procedures and is also secure against rushing adversaries.

Jhanwar and Safavi-Naini [JSN13] consider a model in which all parties (including the adversary) have access to public, shared, randomness. They construct information-theoretically optimal secret sharing schemes in this model by re-encoding Shamir's shares using the available public randomness. This construction achieves the same share length as Shamir's while providing privacy and robustness against any collusion of size less than $n/2$.

Cramer et al. [CDF$^+$08] introduce the notion of *algebraic manipulation detection (AMD) codes*, which is a natural variant of error-detection codes in situations where the adversary's perturbations on a codeword are chosen independently of the codeword. By using this primitive as a pre-code in Shamir's secret sharing scheme (or any secret sharing scheme with linear decoder), they are able to make the scheme robust against adversarial manipulations. The key difference in their model is the notion of robustness; i.e., the requirement is that if the adversary corrupts any of the shares, the reconstruction should *detect* the adversary and fail (rather than output the correct share) with high probability.

More recently, Lewko and Pastro [BP14] defined a variation of robust secret sharing in which the robustness requirement is against *local* adversaries. That is, the error in each share corrupted by the adversary can only depend on the particular share being corrupted. Intuitively, this cor-

responds to the case where a number of adversaries take control of different shares and have to decide on submitting an incorrect share only based on the local information that they possess (the adversaries may agree on a strategy beforehand but cannot communicate after observing their respective shares). They show that even in this restricted model, the minimum required share length is $k + \log(1/\eta) - O(1)$ (under the standard threshold assumption that any set of $t+1$ must reconstruct the secret with probability at least $1-\eta$). Furthermore, they construct efficient schemes in the local model that attains a nearly optimal share length of $k + O(\log(1/\eta))$.

Finally, Safavi-Naini and Wang [SNW15] construct secret sharing schemes based on codes for the wiretap channel problem for the case $n = 2t+1$. This construction is based on wiretap codes that are in turn based on list decodable Reed-Solomon codes, subspace-evasive sets and AMD codes, and attains a share length of $k + O(n^2(\log n)(\log\log n) + n\log(1/\eta))$.

1.2 Our contributions

In this work, we construct an essentially optimal robust secret sharing scheme against possibly adaptive, but non-rushing, adversaries. Somewhat surprisingly, our construction turns out to be strikingly similar to some of the known constructions mentioned in §1.1.

More precisely, the construction first amends the secret with a tag using an AMD code (such as the one in [CDF$^+$08]). Then, it uses Shamir's scheme to encode the result into mn shares,

Ref.	Share length	Efficient?	Remarks
[Sha79]	k	Yes	Only robust against collusions of size $t < n/3$
[JSN13]	k	No	Uses public independent randomness.
[CDF$^+$08]	$k + O(\log(1/\eta))$	Yes	Only robust in the sense of error detection
[BP14]	$k + O(\log(1/\eta))$	Yes	Only secure against local adversaries
[CPS99]	$k + O(n + \log(1/\eta))$	No	
[CFOR12]	$k + \tilde{O}(n + \log(1/\eta))$	Yes	Secure against rushing adversaries
[RBO89]	$k + O(n\log(1/\eta))$	Yes	Secure against rushing adversaries
[SNW15]	$k + \tilde{O}(n^2 + n\log(1/\eta))$	Yes	Restricted to $n = 2t+1$
This work	$k(1+o(1)) + O(\log(1/\eta))$	Yes	Corollary 15
This work	$O_\rho(1)$	Yes	Reconstruction from any $t + \rho n$ shares, for any constant $\rho > 0$, assuming $\frac{t}{n} \le \frac{1}{2} - \rho$, large n and $\eta = \exp(-\Omega(n))$ (Corollary 19).

Table 1: Summary of results in robust secret sharing scheme, and their key features and limitations. The parameter t is the privacy parameter, n is the number of shares and η is the error probability of reconstruction.

for a carefully chosen integer parameter $m > 1$. Finally, the resulting shares are bundled into n groups of size m each which are distributed among the n parties. In other words, we use a variant of Shamir's scheme based on *folded Reed-Solomon codes* (instead of plain Reed-Solomon codes) combined with an AMD pre-code. This is very similar to what used in [CDF+08] to provide robustness in the sense of error-detection, as well as the coding-theoretic construction of Safavi-Naini and Wang [SNW15] (the latter additionally uses subspace-evasive sets that we do not need). Combining Shamir's scheme with some type of information-theoretic pre-code (such as a message authentication code) can also be seen as the underlying idea of other existing constructions such as [CDF01].

The techniques that we use are remarkably simple to describe as well. To prove robustness, we first use an efficient list decoding algorithm of folded Reed-Solomon codes [GR08] to show that the reconstruction procedure always outputs a short list containing an AMD encoding of the correct secret. Second, we use an elegant observation by Guruswami and Smith [GS10] that was used by them to construct "stochastic" error-correcting codes. The observation is that, for any list decodable code that is linear over some base field, the list of potential messages corresponding to the any given received word is the translation of the original message by elements of a set that only depends on the noise vector. In particular, the list of potential messages, shifted by the correct message, is only determined by the code and the error vector chosen by the adversary. For our application in secret sharing, privacy of Shamir's scheme implies that the perturbations of the adversary, and thus the set of error vectors in the message domain, must be independent of the original message and the internal randomness of the AMD code. As a result, the error detection guarantee of the AMD code ensures that, with high probability, all the incorrect potential messages are correctly identified by the reconstruction procedure so that only the correct secret remains in the end.

The above sketch can be made precise to prove our main result as follows.

Theorem 1. *(Corollary 15, rephrased) Let $\delta < 1/2$ be any fixed constant. For any $\eta > 0$, there is an efficient, robust and perfectly private secret sharing scheme with n shares, secret length k, and share length $q \leq k(1 + o(1)) + O(\log(1/\eta))$ that is secure with privacy parameter $t = \delta n$, attaining a reconstruction error of at most η.* □

Same as Shamir's scheme and [SNW15], our result does not necessarily require the observations of the adversary to coincide or overlap with the set of manipulated shares. In fact, the number of adaptive observations by the adversary may in general be different from the number of incorrect shares, and this is allowed as long as the total fraction of observations and incorrect shares add up to a quantity sufficiently smaller than 1.

Although a share length of at least k bits is necessary for any robust secret sharing scheme [Sti92] (even against local, or oblivious, adversaries [BP14]), it is possible to obtain smaller shares at cost of slightly relaxing the threshold property. That is, instead of requiring the secret to be

reconstructible (either with probability 1 or close to 1) from any set of more than t shares, we may require reconstructability from any set of more than $t+g$ shares, for a small "gap" parameter g. A desirable level for the gap parameter is when g is a small fraction of the number of shares, and it is reasonable to argue that a secret sharing scheme that attains such a relaxed threshold property may be of interest to most applications.

We adapt our secret sharing scheme to nonzero gap parameters and, moreover, show that when g is a small fraction of n, the alphabet size may be reduced to an absolute constant (depending on the fraction g/n and assuming that t/n is smaller than $1/2$ by some constant). This is achieved by using folded algebraic geometry codes instead of folded Reed-Solomon codes and their corresponding list decoding algorithms (namely, the state-of-the-art algorithm due to Guruswami and Xing [GX14]). Using algebraic geometry codes, we can prove the following.

Theorem 2. *(Corollary 19, rephrased) For any constant $\rho > 0$, and any $\delta \leq 1/2 - \rho$, there is a constant $q = O_\rho(1)$ such that the following holds. There is a robust and perfectly private secret sharing scheme with n shares, secret length k, and share length $O(q)$, attaining a reconstruction error of $\eta = \exp(-\Omega(\rho n q))$, provided that $n \geq k/(\rho q)$. The scheme satisfies the threshold property in an approximate sense; namely, that the secret can be reconstructed (with probability 1) given any set of $t + \rho n$ shares. The scheme is efficient given polynomial (in n) amount of pre-processed information about the scheme.* □

The efficiency of this scheme is dictated by the efficiency of the underlying list decoding algorithm for algebraic geometry codes. The encoding and list decoding algorithms in [GX14] that we use run in polynomial time provided that a polynomial amount of pre-processed information about the code is available to the algorithms. Naturally, any subsequent improvements in list decoding algorithms of folded algebraic geometry (and for that matter, folded Reed-Solomon) codes would automatically improve the performance of the above secret sharing schemes.

Organization. The rest of the article is organized as follows. We explain the notation in §1.3. Preliminaries, including the exact notion of secret sharing schemes that we use in this work, are discussed in §2. Our general construction is presented and analyzed in §3. We then instantiate the construction using folded Reed-Solomon codes in §4.1 and folded algebraic geometry codes in §4.2. Finally, §4.3 proves optimality of the obtained bounds using a reduction from the wiretap channel problem.

1.3 Notation

We use $d_H(x, y)$ to denote the Hamming distance between two vectors x and y. For a vector $Y = (Y_1, \ldots, Y_n)$, and $i \in [n]$, we use the notation $Y(i)$ to denote Y_i. Moreover, for a sequence $W = (W_1, \ldots, W_t) \in [n]^t$, we use the notation $Y|_W := (Y(W_1), \ldots, Y(W_t))$. All logarithms are to

base two. For a function f and a subset S of the domain of f, we use the notation $f(S)$ to denote the set $\{f(s)\colon s \in S\}$. Moreover for two sets A, B over a group $(\mathcal{G}, +)$, we use $A + B$ to denote $\{a + b\colon a \in A, b \in B\}$, and $A + b$ (for $b \in \mathcal{G}$) to denote $A + \{b\}$.

2 Preliminaries

In this section, we describe the basic notions that are used throughout the paper, including the exact definition of robust secret sharing schemes that we use. The general notion of coding schemes is defined as follows.

Definition 3 (coding scheme)**.** A pair of functions $(\mathsf{Enc}, \mathsf{Dec})$ where $\mathsf{Enc}\colon \mathbb{F}_2^k \times \mathbb{F}_2^\ell \to \mathbb{F}_{2^q}^n$, and $\mathsf{Dec}\colon (\mathbb{F}_{2^q} \cup \{\perp\})^n \to \mathbb{F}_2^k \cup \{\perp\}$ is called a coding scheme if for all $s \in \mathbb{F}_2^k$ and all $z \in \mathbb{F}_2^\ell$, we have $\mathsf{Dec}(\mathsf{Enc}(s, z)) = s$. The function Enc and Dec are respectively called the *encoder* and the *decoder*, and parameters k and q are respectively called the *message length* and the *symbol length*. We use the notation $\mathsf{Enc}(s)$ to denote the random variable $\mathsf{Enc}(s, Z)$ when Z is sampled uniformly at random from $\mathbb{F}_2^\ell$. The coding scheme is called *efficient* if $\mathsf{Enc}, \mathsf{Dec}$ can be computed in polynomial time in nq. The *rate* of the coding scheme is the quantity $k/(nq)$. The coding scheme is binary if $q = 1$.

Using the above definition, we may now define robust secret sharing schemes as a coding scheme satisfying the privacy and robustness requirements.

Definition 4 (robust secret sharing scheme)**.** A robust secret sharing scheme with secret length k, share length q, and number of shares n is a coding scheme $(\mathsf{Share}, \mathsf{Rec})$ with message length k, symbol length q and block length n satisfying the following.

1. **Adaptive privacy:** For a parameter t (known as the *privacy parameter*), and for any "secret" $s \in \mathbb{F}_2^k$, an adversary who (possibly adaptively) observes any up to t of the shares gains (almost or absolutely) no information about the secret s. More formally, for a $Y \in \mathbb{F}_{2^q}^n$, and a parameter t, we define an *observation strategy* as follows. The strategy is specified by an *observation sequence* $W = (W_1, \ldots, W_t)$, where each $W_i \in [n]$ is distinct and determined as a function of $Y(W_1), \ldots, Y(W_{i-1})$. The *observation outcome* with respect to Y is then the string $Y|_W$. The privacy requirement is that for every observation strategy as above, there is a distribution $\mathcal{D}$ over $\mathbb{F}_{2^q}^t$ such that, for every $s \in \mathbb{F}_2^k$, letting $Y := \mathsf{Share}(s)$, the distribution of the observation outcome $Y|_W$ is ϵ-close in statistical distance[1] to $\mathcal{D}$. The scheme satisfies perfect privacy if $\epsilon = 0$.

[1] The statistical distance between two distributions $\mathcal{D}$ and $\mathcal{D}'$ over a finite support Ω is defined as $\mathsf{dist}(\mathcal{D}, \mathcal{D}') := \frac{1}{2}\sum_{x \in \Omega} |\mathcal{D}(x) - \mathcal{D}'(x)|$ and the two distributions are said to be ϵ-close (denoted by $\mathcal{D} \approx_\epsilon \mathcal{D}'$) if $\mathsf{dist}(\mathcal{D}, \mathcal{D}') \leq \epsilon$. In this work, we focus on perfect privacy; i.e., $\epsilon = 0$.

2. **Robustness:** For a parameter d (known as the *robustness parameter*), an adversary who arbitrarily corrupts up to any d of the shares (possibly after adaptively observing any t of the shares) cannot make Rec output an incorrect secret with probability more than η. More formally, consider any observation strategy resulting in an observation sequence W. Then, for any $s \in \mathbb{F}_2^k$ the following must hold. Let $Y := \mathsf{Share}(s)$, and suppose an adversary is given $(W, Y|_W)$ and accordingly chooses an error vector $\Delta \in \mathbb{F}_{2^q}^n$ of Hamming weight at most d. Then it must be that, for some *robustness error* parameter $\eta \geq 0$,

$$\Pr(\mathsf{Rec}(Y + \Delta) \neq s) \leq \eta,$$

where the probability is taken over the internal randomness of Share. The scheme satisfies *perfect robustness* if $\eta = 0$.

The quantity $\log(1/\max\{\eta, \epsilon\})$ is called the *security parameter* of the scheme. We say the scheme satisfies the *threshold property* with gap g if the following holds for all $s \in \mathbb{F}_2^k$ and all sets $S \subseteq [n]$ of size at least $t + g + 1$. Let $Y := \mathsf{Share}(s)$ and $Y' \in (\mathbb{F}_{2^q} \cup \{\perp\})^n$ be so that $Y'|_S = Y|_S$ and $Y'(i) = \perp$ for all $i \in [n] \setminus S$. Then, it must be that

$$\Pr(\mathsf{Rec}(Y') \neq s) \leq \eta,$$

where the probability is taken over the internal randomness of Share. That is, the correct secret can be reconstructed correctly from any set of $t + g + 1$ shares. If $g = 0$, we say that the scheme satisfies a *sharp threshold*. □

An important notion that we use in our constructions is the notion of *algebraic manipulation detection (AMD) codes*, defined as follows.

Definition 5 (AMD code)**.** [CDF+08] A binary coding scheme $(\mathsf{Enc}, \mathsf{Dec})$ with message length k and block length n is an *AMD code* with error η if for every message $s \in \mathbb{F}_2^k$ and every $\Delta \in \mathbb{F}_2^n$, we have

$$\Pr(\mathsf{Dec}(\mathsf{Enc}(s) + \Delta) \notin \{s, \perp\}) \leq \eta,$$

where the probability is taken over the internal randomness of Enc.

The following result is shown in [CDF+08], which we shall use in our constructions. Although, as stated in [CDF+08], the coding scheme is only defined for infinitely many values of the message length k, it can be extended to all integers $k > 0$ by trivial padding techniques without any loss in the asymptotic guarantees.

Theorem 6. *[CDF+08] For every k and parameter $\eta > 0$, there is an efficient AMD code with message length k and encoder of the form*

$$\mathsf{Enc}(s, z) = (s, z, f(s, z))$$

for some $f \colon \mathbb{F}_2^k \times \mathbb{F}_2^q \to \mathbb{F}_2^q$ such that $q = O(\log(1/\eta))$.

The notion of folded codes, following a line of work in algebraic list decoding (originally defined in [GR08]) is the following. Intuitively, a folded code is obtained from an error-correcting code by bundling groups of codeword symbols into "packets" of a certain size, thereby increasing the effective alphabet size in favor of better error resilience guarantees.

Definition 7. Let $\mathcal{C} \subseteq \mathbb{F}_Q^{nm}$ be a code with message length km. The *folded* $\mathcal{C}$ at level m is the code $\mathcal{C}' \subseteq \mathbb{F}_{Q^m}^n$ (with alphabet size Q^m) defined as

$$(c_1, \ldots, c_n) \in \mathcal{C}' \iff ((c_1(1), \ldots, c_1(m)), \ldots, (c_n(1), \ldots, c_n(m)) \in \mathcal{C},$$

where $c_i \in \mathbb{F}_{Q^m}$ and $(c_i(1), \ldots, c_i(m))$ is a natural embedding of $c_i \in \mathbb{F}_{Q^m}$ into $\mathbb{F}_Q^m$. Intuitively, the code $\mathcal{C}$ is obtained by writing each symbol in $\mathcal{C}'$ as a length m vector over $\mathbb{F}_Q$.

3 The construction

The following is the main technical tool used by our constructions, in which we prove that a combination of AMD codes with (folded) linear list decodable codes can be used to construct robust secret sharing schemes.

Theorem 8. *There is a constant $c_0 > 0$ such that the following holds for any integer $k > 0$ and parameter $\eta > 0$. For some $Q = 2^q$ and $m \mid q$, let $\mathcal{C} \subseteq \mathbb{F}_Q^n$ be an explicit $\mathbb{F}_{Q^{1/m}}$-linear code with rate R that is efficiently list decodable from any δ fraction of errors with list size bounded by L and has minimum distance $d > \delta n$. Moreover, suppose $\mathcal{C}$ has a sub-code $\mathcal{C}' \subseteq \mathbb{F}_Q^n$ that, over $\mathbb{F}_{Q^{1/m}}$, is linear with dual distance at least $tm + 1$ and rate $R' \leq R - 1/n$ satisfying*

$$(R - R')nq \geq k + c_0 \log(L/\eta). \tag{1}$$

Then, there is an efficient and perfectly private robust secret sharing scheme $(\mathsf{Share}, \mathsf{Rec})$ *with secret length k and n shares, share length q, privacy parameter t, robustness δn, and robustness error η. Moreover, the scheme satisfies the threshold property with gap $g = n - t - d$.*

Proof. Let $\eta' := \eta/L$. We first instantiate the AMD code of Theorem 6 for message length k and block length

$$n_0 = k + O(\log(1/\eta')) \leq k + c_0(\log(L/\eta))$$

for some constant $c_0 > 0$. Let $(\mathsf{Enc}_0, \mathsf{Dec}_0)$ be the resulting AMD coding scheme.

We can write the code $\mathcal{C}$ as a direct sum $\mathcal{C} = \mathcal{C}' + \mathcal{C}''$ of complementary codes, where $\mathcal{C}'' \subseteq \mathbb{F}_Q^n$ is an $\mathbb{F}_{Q^{1/m}}$-linear sub-code of $\mathcal{C}$ of rate $R - R' > 0$. For the sake of clarity in the sequel we use $\mathcal{C}_0, \mathcal{C}_0' \subseteq (\mathbb{F}_{Q^{1/m}})^{nm}$ to be the codes $\mathcal{C}, \mathcal{C}'$, respectively, when regarded as subspaces of $(\mathbb{F}_{Q^{1/m}})^{nm}$ (in other words, $\mathcal{C}_0, \mathcal{C}_0'$ are the unfolded representations of $\mathcal{C}, \mathcal{C}'$). Recall that $\mathcal{C}_0, \mathcal{C}_0'$ are linear codes over $\mathbb{F}_{Q^{1/m}}$.

Let $f\colon \mathbb{F}_2^{n_0} \to \mathcal{C}''$ be any efficient and $\mathbb{F}_2$-linear invertible function. Such a function exists since $\log_2 |\mathcal{C}''| = (R - R')nq \geq n_0$ by (1). Note that there is also an efficiently computable $\mathbb{F}_2$-linear projection $f'\colon \mathbb{F}_Q^n \to \mathbb{F}_2^{n_0}$ such that for any $w \in \mathcal{C}'$, and any $x \in \mathbb{F}_2^{n_0}$, we have $f'(w + f(x)) = x$.

We define the secret sharing scheme $(\mathsf{Share}, \mathsf{Rec})$ as follows:

- Share: Given $s \in \mathbb{F}_2^k$, $\mathsf{Share}(s)$ first computes $S' := \mathsf{Enc}_0(s)$. Then, it samples a $Z \in \mathbb{F}_Q^n$ according to the uniform distribution on $\mathcal{C}'$ and outputs $Y := f(S') + Z$.
- Rec: Given $Y' \in \mathbb{F}_Q^n$, the procedure $\mathsf{Rec}(Y')$ first uses the list decoding algorithm of $\mathcal{C}$ to compute a list $M \subseteq \mathbb{F}_Q^n$ of size at most L consisting of all codewords of $\mathcal{C}$ that agree with Y' in at least $1 - \delta$ fraction of the positions. Let $M' \subseteq \mathbb{F}_2^{n_0}$ be the set $M' := f'(M)$. If the set $\mathsf{Dec}_0(M') \setminus \{\perp\}$ contains only one element, the algorithm outputs the unique element. Otherwise, the algorithm returns $\perp$.

Let $Y = \mathsf{Share}(s)$ denote the correct shares and $Y' \in \mathbb{F}_Q^n$ be the perturbation of Y according to the strategy of the adversary. Note that Y is always a codeword of $\mathcal{C}$. Furthermore, we are guaranteed that Y' differs from Y in at most δn positions (chosen arbitrarily according to the observation of the adversary). Since the minimum distance of $\mathcal{C}$ is larger than δn and since $\mathcal{C}$ is a linear code, given Y' the decoder can efficiently check whether $Y = Y'$ and make sure that $|M| = 1$ if this is the case, so that there are no ambiguities when no perturbations occur (e.g., using the parity check matrix of $\mathcal{C}$). Since $\mathsf{Dec}_0(\mathsf{Enc}_0(s)) = s$ with probability 1, it follows that $\mathsf{Rec}(\mathsf{Share}(s)) = s$ with probability 1 as well. Therefore, it follows that $(\mathsf{Share}, \mathsf{Rec})$ is indeed a valid coding scheme.

In order to see the privacy requirement, we observe that since $\mathcal{C}'_0$ has dual distance greater than tm and $Z \in \mathbb{F}_Q^n$ is a uniformly random codeword of $\mathcal{C}'$ (and thus, of $\mathcal{C}'_0$ when unfolded), the vector Z is (tm)-wise independent over $(\mathbb{F}_{Q^{1/m}})^{nm}$ (and t-wise independent over $\mathbb{F}_Q^n$). That is, restriction of $Z \in \mathbb{F}_Q^n$ to any t coordinate positions (that may be chosen adaptively) is uniformly distributed on $\mathbb{F}_Q^t$. Therefore, since Z is independent of the randomness of the AMD code, we see that regardless of the message s (and even more generally, conditioned on any particular outcome of S'), the encoding $Y = f(S') + Z$ is t-wise independent. This guarantees that the adversary gains no information about s (and in fact S') by observing any up to t of the shares (note that this is true even if the adversary's strategy may depend on s, see Remark 10 below).

In order to verify the threshold property, we first verify that $n - t - d \geq 0$. In order to see this, note that by the Singleton bound [Rot06, §4.1], and since $\dim \mathcal{C}'_0 < \dim \mathcal{C}_0$, we have $tm + 1 \leq nm - \dim \mathcal{C}'^{\perp}_0 + 1 = \dim \mathcal{C}'_0 + 1 = R'nm + 1 \leq Rnm - m + 1$. Again by the Singleton bound, we have $Rn \leq n - d + 1$, which combined with the previous bound gives $t \leq n - d$. Now, since the minimum distance of $\mathcal{C}$ is d, the vector Y can be uniquely recovered (in fact, with probability 1) from any set of $n - d + 1$ shares. Therefore, since the privacy parameter is t, we obtain a gap of $g = (n - d + 1) - t - 1 = n - d - t$.

Finally, we verify the robustness property. Let the random variable V denote the *view* of the adversary after (possibly adaptively) observing up to t shares. That is, V specifies the sequence of coordinate positions observed by the adversary (possibly adaptively and even given the knowledge of s) and the value of shares at each one of those positions. In the sequel, we consider the conditional probability space in which V attains a specific value v; i.e., we condition all random variables on $V = v$. Our goal is to show that under any such conditioning, the robustness guarantee is satisfied. Observe that because of the privacy argument, the two random variables V and S' (where we recall that $S' = \mathsf{Enc}_0(s)$ via the AMD code) are independent. Therefore, the distribution of S' remains unchanged under the conditioning $V = v$.

Now suppose given the observation $V = v$ (and possibly the secret s), the adversary picks a fixed error vector $\Delta \in \mathbb{F}_Q^n$ of Hamming weight at most δn and perturbs Y to $Y' = Y + \Delta$ (if the adversary picks Δ according to a randomized function of v, we may use the following argument for any fixing of the internal randomness of the adversary; i.e., we may add the adversary's randomness to the conditioning).

We now follow an argument similar to Guruswami and Smith [GS10] to complete the robustness analysis. Let $M_{Y,\Delta}$ denote the set of all codewords of $\mathcal{C}$ that differ from Y' in at at most δn coordinate positions. That is,

$$\begin{aligned} M_{Y,\Delta} &:= \{c \in \mathcal{C} \colon d_H(c, Y + \Delta) \leq \delta n\} \\ &= \{c \in \mathcal{C} \colon d_H(Y + c, \Delta) \leq \delta n\} \\ &= Y + \{c \in \mathcal{C} \colon d_H(c, \Delta) \leq \delta n\}, \end{aligned} \tag{2}$$

where the last equality is due to the linearity of the code $\mathcal{C}$. Recall that $S' = f'(Y)$ where f' is an $\mathbb{F}_2$-linear projection function. Now, we apply f' on every element of $M_{Y,\Delta}$ to obtain the set $M' \subseteq \mathbb{F}_2^{n_0}$ that using (2) can be written as follows.

$$\begin{aligned} M' &:= f'(M_{Y,\Delta}) \\ &= f'(Y) + \{f'(c) \colon c \in \mathcal{C} \wedge d_H(c, \Delta) \leq \delta n\} \\ &= S' + \{f'(c) \colon c \in \mathcal{C} \wedge d_H(c, \Delta) \leq \delta n\}. \end{aligned}$$

Observe that, by the above derivation, the set $S' + M'$ is completely determined by the code $\mathcal{C}$ and the fixed shift vector Δ and is otherwise independent of Y and, importantly, the internal randomness of the AMD encoder Enc_0.

Recall that the reconstruction function Rec applies Dec_0 on all elements of M' and outputs a unique valid decoding if it exists (and otherwise, outputs $\perp$). In other words, reconstruction is successful if and only if $|\mathsf{Dec}_0(M') \setminus \{\perp\}| = 1$ (observe that it is already guaranteed that $S' \in M'$ according to list deocdability of $\mathcal{C}$ which ensures that the correct codeword is always on the list).

Let $\Delta' \in S' + M'$ be any shift vector according to M'. Observe that

$$\Pr(\mathsf{Dec}_0(S' + \Delta') \notin \{S', \perp\}) \leq \eta' \tag{3}$$

from the definition of AMD codes. Here, the probability is taken under the conditioning $V = v$, which we have shown to not affect the internal randomness of the AMD encoder (i.e., the distribution of S' remains unchanged under the conditioning $V = v$). Therefore, by a union bound,

$$\Pr(|\mathsf{Dec}_0(M') \setminus \{\perp\}| \neq 1) \leq |M'|\eta' \leq L\eta' = \eta,$$

which concludes the robustness analysis. Observe that (3) in fact shows that, with probability at least $1-\eta$, the decoder never outputs any secret other than the correct s (if it does not output $\perp$). This is an interesting property as pointed out in Remark 11 below.

□

Remark 9. The minimum distance bound $d > \delta n$ in Theorem 8 is only used to make sure that the scheme $(\mathsf{Share}, \mathsf{Rec})$ is a valid coding scheme; i.e., that $\Pr(\mathsf{Rec}(\mathsf{Share}(s)) = s) = 1$. If instead one wishes to have $\Pr(\mathsf{Rec}(\mathsf{Share}(s)) = s) \geq 1 - \eta$ (or if $\mathcal{C}$ has a decoder that produces a list of size 1 given a correct codeword), this requirement can be eliminated.

Remark 10. As mentioned in the proof of Theorem 8, the theorem holds even if the adversary's observation and perturbation strategies depend on the secret s. This is a property that also holds true for the original Shamir's scheme.

Remark 11. As observed in the end of the proof of Theorem 8, the secret sharing scheme of Theorem 8 *never* outputs a wrong secret (except with probability at most η). That is, even if the fraction of adversarial perturbations exceeds the designated parameter δ and the scheme fails to reconstruct the correct share s, it has to fail with probability at least $1 - \eta$. This is an appealing property that is not present in Shamir's original secret sharing scheme.

4 Instantiations

4.1 Construction based on Reed-Solomon codes

In this section, we instantiate Theorem 8 using folded Reed-Solomon codes. When folding (Definition 7) is instantiated to the special case of Reed-Solomon codes, we have the following definition of folded Reed-Solomon codes.

Definition 12. Let q be a prime power. A *folded Reed-Solomon* code with block length n, alphabet size Q^m and message length k can be specified as the image of an encoder $\mathsf{Enc}\colon (\mathbb{F}_Q^m)^k \to (\mathbb{F}_Q^m)^n$ where $\mathsf{Enc}(f)$ interprets the input f as a polynomial of degree $mk-1$ over $\mathbb{F}_Q$ and outputs a vector $(F_1, \ldots, F_n)$ (where $F_i \in \mathbb{F}_Q^m$) such that $F_i = (f(\alpha_{i,1}), \ldots, f(\alpha_{i,m}))$ and the sequence $(\alpha_{i,j} \colon i \in [n], j \in [m])$ is a sequence of distinct evaluation points over $\mathbb{F}_Q$ explicitly specified by the code design. Rate of the folded Reed-Solomon code is k/n, and the code is linear over $\mathbb{F}_Q$.

As shown in [GR08], folded Reed-Solomon codes attain an optimal trade-off between rate and list decoding radius. Specifically, the following is the main result proved[2] in [GR08].

Theorem 13. *[GR08, follows from Theorem 4.3] For any prime power p, integers $n > k > 0$ and integer $c \geq 1$ and constant parameter $\rho > 0$, there is an $\mathbb{F}_p$-linear folded Reed-Solomon code with message length k and block length n such that for some $L = n^{O(\log^2(1/\rho)/\rho^2)}$ and $\delta \geq 1 - k/n - \rho$, the following hold: (1) The code is list decodable from any δ fraction of errors with list size at most L; (2) The alphabet size of the code is equal to L^c; (3) The code is linear over $\mathbb{F}_p$.*

We now apply the above result in Theorem 8 to obtain the main result of this section, as follows.

Theorem 14. *For every integers $n > t \geq 1$, $g \geq 0$ and real parameters $\delta, \nu, \eta > 0$ such that*

$$\rho := 1 - \delta - \frac{t + g + 1}{n} > 0$$

there is a $q_0 = O(\frac{\log^2(1/\rho)}{\nu\rho^2} \log n)$ such that for any integer $q \geq q_0$ the following holds. There is an efficient and perfectly private secret sharing scheme $(\mathsf{Share}, \mathsf{Rec})$ with n shares, share length q, privacy parameter t, threshold property with gap g, and secret length k satisfying $k \geq (1 + g - \nu)q - O(\log(1/\eta))$. Moreover, the scheme achieves a robustness parameter of δn and robustness error η.

Proof. Let $c := \lfloor c_0/\nu \rfloor$, where c_0 is the constant from Theorem 8. Let $\mathcal{C} \subseteq \mathbb{F}_Q^n$ be an $\mathbb{F}_2$-linear folded Reed-Solomon code, as obtained by Theorem 13, of length n, message length $k' := t + g + 1$, rate $R := k'/n$, alphabet size $Q = L^c$ for some $L = n^{O(\log^2(1/\rho)/\rho^2)}$ that is list decodable from any $\delta = 1 - R - \rho$ fraction of errors with list size bounded by L.

We instantiate Theorem 8 with the code $\mathcal{C}$ to obtain a secret sharing scheme $(\mathsf{Share}, \mathsf{Rec})$ with share length $q := \log Q = c \log L = O(\frac{\log^2(1/\rho)}{\nu\rho^2} \log n)$. We now verify that the requirements of Theorem 8 are satisfied.

First, note that since any folded Reed-Solomon code is on the Singleton bound, the distance d of $\mathcal{C}$ satisfies $d = n - k' + 1 = (1 - R)n + 1 > (\delta + \rho)n > \delta n$.

Let $\mathsf{Enc}_{\mathcal{C}} \colon \mathbb{F}_Q^{k'} \to \mathbb{F}_Q^n$ be the natural encoder for the code $\mathcal{C}$. That is, $\mathsf{Enc}_{\mathcal{C}}$ interprets the input as a univariate polynomial f of degree $k'm - 1$ over a subfield of $\mathbb{F}_Q$ of size $Q^{1/m}$, for some integer $m > 0$, and evaluates f at nm points, interpreting the result as n points over $\mathbb{F}_Q$, each consisting of a bundle of m evaluations (cf. Definition 12). We set the sub-code $\mathcal{C}'$ needed by Theorem 8 to be the code obtained by setting the last $k' - t$ (among the total of k') of the inputs of $\mathsf{Enc}_{\mathcal{C}}$ to be zeros (in algebraic terms, we take the subcode $\mathcal{C}'$ to be the folded Reed-Solomon code formed by the space of univariate polynomials, over $\mathbb{F}_{Q^{1/m}}$, of degree at most $tm - 1$). Thus, the subcode $\mathcal{C}'$ (as a code over

[2] As stated in [GR08], the result is not shown for all choices of the block length n. However, trivially one can obtain a family of codes for all block lengths by adding additional evaluation points that are not used by the decoder, without incurring an adverse effect in the asymptotic bounds.

$\mathbb{F}_{Q^{1/m}}$) is a Reed-Solomon code of dimension tm and dual distance $nm-(nm-tm)+1 = tm+1$. Moreover, the rate R' of $\mathcal{C}'$ is equal to $t/n = (k'-g-1)/n \leq R - 1/n$, and we thus have

$$(R-R')nq = (k'-t)q = (g+1)q,$$

which, by (1) in the statement of Theorem 8, we wish to be at least

$$k + c_0 \log(L/\eta) = kc_0 \log(1/\eta) + \frac{c_0}{c} q \geq kc_0 \log(1/\eta) + \nu q.$$

Therefore, by choosing k to be the largest integer satisfying the above, we may ensure that $k \geq (g+1-\nu)q - O(\log(1/\eta))$ as desired.

Finally, to verify the threshold property, by Theorem 8 we have that the gap achieved by the code is upper bounded by $n - t - d = n - t - (n - k' + 1) = g$.

□

For the important special case of $\delta = t/n$ and $g = 0$ we derive the following immediate corollary from Theorem 14.

Corollary 15. *Let $\delta < 1/2$ be any fixed constant. For every integer $n > 1/(1-2\delta)$ and parameters $\eta > 0$ and $\nu > 0$, there is an efficient and perfectly private secret sharing scheme* (Share, Rec) *with n shares, share length $q = O_\nu(\log n)$, and secret length $k \geq q(1-\nu) - O(\log(1/\eta))$. The scheme attains a sharp threshold, privacy and robustness δn, and robustness error η.* □

4.2 Reducing the share length using algebraic geometry codes

A slight drawback of the result in Corollary 15 is that the share length grows with the number of shares (i.e., $q \to \infty$ as $n \to \infty$). This is a direct consequence of the fact that the alphabet size of a Reed-Solomon must grow with its block length. In order to resolve this issue, we instantiate Theorem 8 with a family of folded algebraic geometry (AG) codes as described in [GX14]. As we see in this section, for any fixed $\delta < 1/2$, this results in a secret sharing scheme with privacy and robustness δn and constant alphabet size (depending on $1 - 2\delta$).

Theorem 16. *[GX14, Theoren 4.3] For any $\rho > 0$ and a real $R \in (0,1)$, one can construct a folded algebraic geometry code over alphabet size $Q = (1/\rho)^{O(1/\rho^2)}$ with rate at least R and decoding radius $\delta = 1 - R - \rho$ such that the length n of the code tends to infinity and is independent of ρ. Moreover, the code is deterministically list decodable with a list size $O(n^{1/\rho^2})$. Given a polynomial (in n) amount of pre-processed information about the code, the algorithm runs in deterministic polynomial time.*

Theorem 17. *Let c_0 be the constant from Theorem 8. For any constants $\rho, \delta > 0$, there is an integer $q = \Theta(\log(1/\rho)/\rho^2)$ and $n_0 = (1/\rho)^{O(1)}$ such that for all integers t, k and $n \geq n_0$ and real*

parameter $\eta > 0$ that satisfy

$$\frac{k}{qn} + \frac{t}{n} + \delta \leq 1 - \rho - c_0 \frac{\log(1/\eta)}{nq} \tag{4}$$

the following holds. There is an efficient and perfectly private secret sharing scheme (Share, Rec) *with n shares, share length q, privacy parameter t and secret length k. Moreover, the scheme achieves a robustness parameter of δn and error η, and satisfies the threshold property with gap at most $n(1 - \frac{t}{n} - \delta)$. The scheme is efficient given polynomial (in n) amount of pre-processed information about the scheme.*

Proof. The proof is similar to that of Theorem 14, but uses the folded algebraic geometry codes of Theorem 16 instead of folded Reed-Solomon codes.

Let $\rho' = \Theta(\rho)$ to be a parameter to be determined later. Let $\mathcal{C}$ be a folded algebraic geometry code of length[3] n and rate $R = 1 - \delta - \rho'$ over alphabet size $Q = (1/\rho)^{\Theta(1/\rho^2)}$ that is list decodable from any δ fraction of errors with list size $L = O(n^{1/\rho'^2})$. Let $k' := Rn$ be the message length of $\mathcal{C}$. We apply Theorem 8 on this code to obtain a secret sharing scheme (Share, Rec) with n shares of length $q = \log Q = \Theta(\log(1/\rho)/\rho^2)$. Now we set up the parameters so as to satisfy the requirements of Theorem 8.

We observe that the construction of Theorem 16 uses function fields over Garcia-Stichtenoth towers, and the setup of the parameters is so that the genus G of the function field can be made to be at most $\rho' nm$, where m is the depth of folding, or in other words, nm is the block length of the code before folding. Therefore, by the Riemann-Roch Theorem ([Sti09, Theorem 1.5.15 combined with Corollary 2.2.3]), the minimum distance of $\mathcal{C}$ is greater than $n - k' - G/m \geq n - k' - \rho' n = n(1 - R - \rho') = \delta n$.

Let $\mathcal{C}_0 \subseteq (\mathbb{F}_{Q^{1/m}})^{nm}$ to be the unfolded representation of $\mathcal{C}$ (thus $\mathcal{C}_0$ is the original, unfolded, algebraic geometry code). As is the case with Reed-Solomon codes, one can identify a subcode $\mathcal{C}' \subsetneq \mathcal{C}_0$, over the same function field as $\mathcal{C}_0$, of dimension $t' := tm + \lceil 2\rho' nm \rceil + 4$ over $\mathbb{F}_{Q^{1/m}}$. Let R' be the rate of $\mathcal{C}'$. We will have $R' \leq R - 1/n$ assuming that $t' \leq (k' - 1)m$. The dual of $\mathcal{C}'$ has dimension $nm - \dim(\mathcal{C}') = nm - t'$ and, by [Sti09, Theorem 2.2.7 combined with Corollary 2.2.3 and Proposition 2.1.8], minimum distance at least

$$\dim(\mathcal{C}') - 2G - 3 = t' - 2G - 3 \geq t' - 2\rho nm - 3 > tm.$$

In order to satisfy (1), noting that

$$(R - R')nq \geq (1 - \delta - \rho)nq - t'q/m \geq nq\big(1 - \delta - \rho' - \frac{tm + 2\rho' nm + 5}{mn}\big) \geq nq(1 - \delta - \frac{t}{n} - 3\rho' - 5),$$

[3]Even though Theorem 16 constructs codes for infinitely many choices of n, without loss of generality one can assume that there is a code for every n. Since the set of block lengths for which the family contains a code is sufficiently dense, this can be ensured by trivial padding without any loss in the asymptotic parameters.

it suffices to ensure that

$$\frac{k}{nq} + \frac{t}{n} + \delta + \frac{c_0 \log(1/\eta)}{nq} \leq 1 - 3\rho' - \frac{5 + c_0 \log L}{nq}. \tag{5}$$

Recall that $\log L \leq (1/\rho'^2) \log n + O(1)$. Thus by choosing an appropriate $n_0 = (1/\rho)^{O(1)}$ and ensuring that $n \geq n_0$, and $\rho' \leq \rho/4$ we can make the right hand side of (5) at least $1 - \rho$. Consequently, assuming (4), i.e.,

$$\frac{k + c_0 \log(1/\eta)}{nq} + \frac{t}{n} + \delta \leq 1 - \rho,$$

we have (5) and, in turn, (1).

Finally, by Theorem 8, the scheme satisfies the threshold property with gap $g = n - t - d \leq n(1 - \frac{t}{n} - \delta)$, as desired. □

From this result, we obtain the following corollary.

Corollary 18. *For any constants $\delta, \gamma, \rho > 0$, there is a $q_0 = O(\log(1/\rho)/\rho^2)$ and $n_0 = O(1/\rho)$ such that for all integers $c \geq 1$, the following holds. Let $q := cq_0$. For any integers $k > 0$, $n \geq n_0$, and parameter $\eta > 0$ such that*

$$\frac{k}{qn} + \gamma + \delta \leq 1 - \rho, \tag{6}$$

There is a perfectly private secret sharing scheme $(\mathsf{Share}, \mathsf{Rec})$ with n shares, secret length k, share length q, privacy parameter at least γn, and threshold property with gap at most $n(1 - \delta - \gamma)$. Moreover, the scheme achieves a robustness parameter of δn and error $\eta = \exp(-\Omega(\rho n q))$. The scheme is efficient given polynomial (in n) amount of pre-processed information about the scheme.

Proof. We simply apply Theorem 17 with constant $\rho' := \rho/2$ (for the parameter ρ required by Theorem 17) to obtain a secret sharing scheme $(\mathsf{Share}, \mathsf{Rec})$ with cn shares, secret length k, share length $q_0 = O(\log(1/\rho)/\rho^2)$, robustness δcn, and privacy parameter $t := \lceil \gamma cn \rceil$.

Let c_0 be the constant from Theorem 8. We choose the error parameter $\eta = \exp(-\Omega(\rho n q))$ so that $c_0 \log(1/\eta) \leq \rho n q/2 - 1$, and thus

$$\frac{k}{qn} + \frac{t}{cn} + \delta + c_0 \frac{\log(1/\eta)}{nq} \leq 1 - \rho'$$

as needed by Theorem 17. Next, we bundle disjoint groups of c shares into shares of length $cq_0 = q$, thus obtaining a scheme with n shares of length q and the desired parameters. □

Corollary 18, in turn, immediately implies the following result on robust secret sharing with privacy and robustness parameter δn for any $\delta < 1/2$.

Corollary 19. *For any constant $\rho > 0$, and any $\delta \leq 1/2 - \rho$, There is a $q_0 = O(\log(1/\rho)/\rho^2)$ such that for any $q \geq q_0$ and integers $k > 0$ and $n \geq k/(\rho q)$, the following holds. There is a perfectly private secret sharing scheme* $(\mathsf{Share}, \mathsf{Rec})$ *with n shares, secret length k, and share length at most $2q$. The scheme attains privacy and robustness parameters equal to δn and error $\eta = \exp(-\Omega(\rho n q))$, and satisfies the threshold property with gap at most $2\rho n$. The scheme is efficient given polynomial (in n) amount of pre-processed information about the scheme.* □

Compared with the result of Corollary 15 obtained from Reed-Solomon codes, we see that the share length q can be chosen to be a constant (depending on the difference $1/2 - \delta$), and at the same time the number of shares can be made arbitrarily large as well. However, for this to be possible when the designed share length is small, the number of shares n needs to be large enough[4] so that $n \geq k/(\rho q)$. In Section 4.3 we show that this is necessary for any robust secret sharing scheme with share length q that attains privacy and robustness parameters close to $n/2$.

4.3 Optimality

In this section we briefly demonstrate that, for a general share length q, a robust secret sharing scheme satisfying (6) for arbitrarily small $\rho > 0$ is essentially optimal (even if the threshold property is not a concern). This can be shown by a straightforward reduction from the *wiretap channel problem*.

In the wiretap channel problem [Wyn75, CK78], the goal is to construct a coding scheme to encode a secret $S \in \mathbb{F}_2^k$ to an encoding $Y \in \mathbb{F}_Q^n$ that is transmitted over a *main* channel to a recipient. The encoding is additionally sent to an adversary over a *wiretap* channel that has a smaller channel capacity compared to the main channel. The secrecy requirement of the problem is that the adversary should not learn any information about the secret from the wiretap channel's observation, whereas the recipient observing the main channel should be able to reconstruct the correct secret (with probability at least $1 - \eta$ for arbitrarily small $\eta > 0$). There are various formulations of the problem that differ in the following aspects:

1. Whether the reconstruction and secrecy requirements are defined with respect to a uniformly random secret S or, more stringently, the worst case secret,

2. The choice of the main and wiretap channels, and

3. The notion of secrecy. In weak secrecy, the requirement is the mutual information security (cf. [BTV12]) of the form
$$I(S; Y') \leq \epsilon k,$$

[4] Note such a requirement is not a barrier for the Reed-Solomon based constructions such as Shamir's scheme and the result of Theorem 14, since we have $q \geq k$ in those schemes.

where Y' is the wiretap channel's output, for arbitrarily small $\epsilon > 0$. A much stronger notion is *semantic security* (formalized in [BTV12]) which requires that there must be a distribution $\mathcal{D}$, determined by the coding scheme, such that for every fixed secret $s \in \mathbb{F}_2^k$, the wiretap channel's output is statistically ϵ-close to $\mathcal{D}$.

An important parameter to characterize is the *secrecy capacity* in this model, which is the highest achievable rate $R := k/(qn)$ by a coding scheme satisfying the above-mentioned reconstruction and secrecy requirements. For our reduction, the main channel is the Q-ary erasure channel, where $Q := 2^q$, with erasure probability p and, moreover, the wiretap channel is the Q-ary symmetric channel with error probability p', for given parameters p and p'. In this case, it is known that the secrecy capacity even with respect to a random secret and weak secrecy requirement is the difference between capacities of the two channels [CK78, LYC77], which is equal to $(1 - h_Q(p')) - (1 - p) = p - h_Q(p') \leq p - p'$, where $h_Q(\cdot)$ is the Q-ary entropy function.

It is immediate that a robust secret sharing scheme (as formulated in Definition 4) satisfies the requirements of the wiretap channel problem formulated above, provided that the robustness parameters is set to be $\delta n := (p' + \rho')n$, for an arbitrarily small $\rho' > 0$, and the privacy parameter is set to be $t := \lceil (1 - p + \rho')n \rceil$.

In fact a secret sharing scheme is a stronger object than needed since it allows for the erasure positions and also perturbations to be adaptively chosen by the adversary. Moreover, it provides secrecy for worst-case secrets as well as semantic security (in fact, recall that our constructions achieve perfect secrecy; i.e., semantic security with $\epsilon = 0$).

By Chernoff bounds, the probability η' that the fraction of erasures for the adversary is less than $p - \rho'$ or the fraction of perturbations in the direct channel is more than $p' + \rho'$ is exponentially small (i.e., at most $\eta' = \exp(-\Omega(n))$ for any $\rho' > 0$ that is a constant). It follows that the correctness requirement of the wiretap channel problem can be satisfied with error at most $\eta + \eta' = o(1)$ (provided that $\eta = o(1)$) and, moreover, semantic secrecy is also satisfied with a statistical error of $\epsilon \leq \eta'$ (where the choice of $\mathcal{D}$ would be the uniform distribution over $\mathbb{F}_Q^t$).

Since the secrecy capacity of the above wiretap channel problem is at most $p - p'$, it must be that, defining $\gamma := t/n$,

$$\frac{k}{qn} \leq p - p' \leq 1 - \gamma - \delta + (2\rho' + o(1)).$$

Thus the bound obtained in (6) is the best to hope for.

References

[Bla79] G.R. Blakley. Safeguarding cryptographic keys. In *National Computer Conference*, volume 48, pages 313–317. Springer, 1979.

[BP14] A. Bishop Lewko and V. Pastro. Robust secret sharing schemes against local adversaries. Cryptology ePrint Archive, Report 2014/909, 2014. `http://eprint.iacr.org/`.

[BTV12] M. Bellare, S. Tessaro, and A. Vardy. Semantic security for the wiretap channel. In *Proceedings of Advances in Cryptology CRYPTO 2012*, volume 7417 of *Lecture Notes in Computer Science*, pages 294–311. Springer, 2012.

[CDF01] R. Cramer, I. Damgård, and S. Fehr. On the cost of reconstructing a secret, or VSS with optimal reconstruction phase. In *Proceedings of Advances in Cryptology CRYPTO 2001*, volume 2139 of *Lecture Notes in Computer Science*, pages 503–523. Springer, 2001.

[CDF+08] R. Cramer, Y. Dodis, S. Fehr, C. Padró, and D. Wichs. Detection of algebraic manipulation with applications to robust secret sharing and fuzzy extractors. In *Advances in Cryptology - EUROCRYPT 2008*, volume 4965 of *Lecture Notes in Computer Science*, pages 471–488. Springer, 2008.

[CFOR12] A. Cevallos, S. Fehr, R. Ostrovsky, and Y. Rabani. Unconditionally-secure robust secret sharing with compact shares. In *Proceedings of Advances in Cryptology EUROCRYPT 2012*, volume 7237 of *Lecture Notes in Computer Science*, pages 195–208. Springer, 2012.

[CK78] I. Csiszár and J. Körner. Broadcast channels with confidential messages. *IEEE Transactions on Information Theory*, 24(3):339–348, 1978.

[CPS99] S. Cabello, C. Padró, and G. Sáez. Secret sharing schemes with detection of cheaters for a general access structure. In *Proceedings of Fundamentals of Computation Theory*, volume 1684 of *Lecture Notes in Computer Science*, pages 185–194. Springer, 1999.

[GR08] V. Guruswami and A. Rudra. Explicit codes achieving list decoding capacity: Error-correction with optimal redundancy. *IEEE Transactions on Information Theory*, 54(1):135–150, 2008.

[GS10] V. Guruswami and A. Smith. Codes for computationally simple channels: Explicit constructions with optimal rate. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS 2010)*, pages 723–732, 2010.

[GX14] V. Guruswami and C. Xing. *Optimal rate list decoding of folded algebraic-geometric codes over constant-sized alphabets*, pages 1858–1866. 2014.

[IOS12] Y. Ishai, R. Ostrovsky, and H. Seyalioglu. Identifying cheaters without an honest majority. In *Proceedings of Theory of Cryptography (TCC 2012)*, volume 7194 of *Lecture Notes in Computer Science*, pages 21–38. Springer, 2012.

[JSN13] M. Jhanwar and R. Safavi-Naini. Unconditionally-secure robust secret sharing with minimum share size. In *Proceedings of Financial Cryptography and Data Security*, volume 7859 of *Lecture Notes in Computer Science*, pages 96–110. Springer, 2013.

[LYC77] S. Leung-Yan-Cheong. On a special class of wiretap channels (corresp.). *IEEE Transactions on Information Theory*, 23(5):625–627, 1977.

[RBO89] T. Rabin and M. Ben-Or. Verifiable secret sharing and multiparty protocols with honest majority. In *Proceedings of the Twenty-first Annual ACM Symposium on Theory of Computing (STOC '89)*, pages 73–85, 1989.

[Rot06] R.M. Roth. *Introduction to Coding Theory*. Cambridge University Press, 2006.

[Sha79] A. Shamir. How to share a secret. *Communications of the ACM*, 22(11):612–613, 1979.

[SNW15] R. Safavi-Naini and P. Wang. A model for adversarial wiretap channels and its applications. *Journal of Information Processing*, 23(5):554–561, 2015.

[Sti92] D.R. Stinson. An explication of secret sharing schemes. *Designs, Codes and Cryptography*, 2(4):357–390, 1992.

[Sti09] H. Stichtenoth. *Algebraic Function Fields and Codes*. Springer, second edition, 2009.

[Wyn75] A. D. Wyner. The wire-tap channel. *The Bell System Technical Journal*, 54:1355–1387, 1975.