Designs, Codes and Cryptography (2019) 87:1777–1796
https://doi.org/10.1007/s10623-018-0578-y

Nearly optimal robust secret sharing

Mahdi Cheraghchi[1]

Received: 23 August 2017 / Revised: 14 July 2018 / Accepted: 25 October 2018 /
Published online: 1 November 2018

Abstract
We prove that a known general 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 $\delta \in (0, 1/2)$. Shamir's original scheme is robust for all $\delta \in (0, 1/3)$. Beyond that, we employ the best known list decoding algorithms for Reed-Solomon codes and show that, with high probability, only the correct secret maintains the correct information-theoretic tag if an algebraic manipulation detection (AMD) code is used to tag secrets. This result holds in the so-called "non-rushing" model in which the n shares are submitted simultaneously for reconstruction. We thus obtain a fully explicit 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, the share length can be decreased 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. From a practical perspective, the main importance of our result is in showing that existing systems employing Shamir-type secret sharing schemes can be made much more robust than previously thought with minimal change, essentially only involving the addition of a short and simple checksum to the original data.

Keywords Coding and information theory · Cryptography · Algebraic coding theory

Mathematics Subject Classification 68P30 · 94A60 · 11T71

Communicated by M. Paterson.

A preliminary version of this work appears in Proceedings of the IEEE International Symposium on Information Theory (ISIT 2016).

✉ Mahdi Cheraghchi
m.cheraghchi@imperial.ac.uk

[1] Department of Computing, Imperial College London, London, UK

1 Introduction

Secret sharing, introduced by the seminal works of Shamir [23] and Blakley [1], 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.

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 at most 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. [22, Chap. 6]). As a result, a straightforward calculation shows that Shamir's secret sharing scheme is *robust*, in the sense that it can tolerate malicious parties that submit incorrect shares. In particular, the correct secret s can be always reconstructed even if up to a third 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., [24]).

Table 1 Summary of results in robust secret sharing scheme, and their key features and limitations

Ref.	Share length	Efficient?	Remarks
[23]	k	Yes	Only robust against collusions of size $t < n/3$
[8]	$k + O(\log(1/\eta))$	Yes	Only robust in the sense of error detection
[2]	$k + O(\log(1/\eta))$	Yes	Only secure against local adversaries
[12]	$k + O(n + \log(1/\eta))$	No	
[10]	$k + \tilde{O}(n + \log(1/\eta))$	Yes	Secure against rushing adversaries
[21]	$k + O(n \log(1/\eta))$	Yes	Secure against rushing adversaries
[24]	$k + \tilde{O}(n^2 + n \log(1/\eta))$	Yes	For $n = 2t + 1$
[6]	$O(1)$	Yes	Monte-Carlo, reconstruction from $t + \Omega(n)$ of the shares, for large n, and $\eta = \exp(-\Omega(n))$.
[3]	$k + O(\log(1/\eta)(\log^4 n + (\log^3 n) \log k))$	Yes	For $n = 2t + 1$
This work	$k(1 + o(1)) + O(\log(1/\eta))$	Yes	Corollary 14
This work	$O(1)$	Yes	Corollary 18
			Reconstruction from any $t + \rho n$ shares, for any constant $\rho > 0$, assuming $\frac{t}{n} \leq \frac{1}{2} - \rho$, large n and $\eta = \exp(-\Omega(n))$ (Corollary 18).

The parameter t is the privacy parameter, n is the number of shares and η is the error probability of reconstruction

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 $1/2$; i.e., when $n \leq 2t$ [19]. 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 [21] 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 [25].

Another scheme, due to Cramer et al. [7] (and based on [12] 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. [10]).

Cevallos et al. [10] propose a scheme similar to [21] 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.

Cramer et al. [8] 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 [2] 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 corresponds 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 $k + O(\log(1/\eta))$.

In another recent work, Cramer et al. [6] combine AMD codes with universal hash functions and (folded) list decodable codes to construct a secret sharing scheme with potentially constant share length (more precisely, share length $\Theta(1 + \log(1/\eta)/n)$). Their construction is with respect to a randomly chosen hash function from a universal family and is thus a Monte-Carlo construction. That is, the code construction relies on the probabilistic method (and thus may not result in the desired secret sharing scheme with unfortunate choices of the randomness), however the encoder and decoders are efficient once the randomness of the code construction is set to an appropriate choice. Moreover, this construction considers the "ramp model" in which it is not necessary to be able to reconstruct the secret from any $t + 1$ of the shares. This relaxation is in fact necessary for any secret sharing scheme with share length smaller than the secret length k.

Finally, Safavi-Naini and Wang [24] 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 $k + O(n^2(\log n)(\log\log n) + n\log(1/\eta))$.

Subsequent to a preliminary draft of the present work, Bishop et al. [3] construct an efficient and nearly optimal robust secret sharing scheme for $n = 2t + 1$ that achieves share length $k + O(\log(1/\eta)(\log^4 n + (\log^3 n)\log k))$. In general, this is incomparable with the bound we achieve (while both being very close to the optimal $k + O(\log(1/\eta))$. This work follows the authentication graph idea of Rabin and Ben-Or [21] (in which a MAC signature is used for every party in each share to authenticate the shares for every other party) and its improvement by Cevallos et al. [10]. In particular, [3] considers a subsampled authentication graph, leading to nearly optimal share lengths, which is then shown to provide robustness via a delicate analysis based on the approximation algorithms for the minimum graph bisection problem. It is, however, not shown whether this improvement maintains robustness against rushing adversaries.

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 Sect. 1.1 and involves a simple modification of Shamir's original secret sharing scheme.

More precisely, the construction first amends the secret with a tag using an AMD code (such as the one in [8]). Then, it uses Shamir's scheme to encode the result into mn shares, 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 [8] to provide robustness in the sense of error-detection, as well as the coding-theoretic construction of Safavi-Naini and Wang [24] (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 [7].

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 [15] 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 [16] 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 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 at the end.

Our construction and underlying ideas share an overlap with the above-mentioned recent result of Cramer et al. [6] in which the authors construct a Monte-Carlo secret sharing scheme with small share length in the ramp model (where obtaining a sharp threshold; i.e., reconstructability from any $t + 1$ shares, is not a requirement[1]). The construction in that work can be described as follows: First, the secret s is encoded with an AMD code, and then the result x is mapped to a random element in $h^{-1}(x)$, where h is a *fixed* and appropriately chosen linear hash function. The resulting sequence is finally encoded using a list decodable code. Unfortunately, this result does not determine an explicit suitable choice for h. However it shows, using the probabilistic method, that most functions in a universal family of hash functions are suitable choices for the hash function h. In other words, if h is randomly[2] picked from a universal family of hash functions, with high probability over the choice of h the resulting scheme is robust with the desired parameters. Therefore, the hash function h is determined by the *code construction* once and for all, and the probabilistic method shows that most choices of h would result in equally good secret sharing schemes. It is not clear, however, whether one can efficiently and deterministically find a suitable choice for the hash function h without running an exponential-time computation (as is usual in random coding arguments). Compared with this result, our work completely eliminates the need for the hash

[1] It should however be noted that any (robust) ramp secret sharing scheme can be modified to also satisfy a sharp thereshold by simply adding Shamir shares to each existing share, at cost of increasing the share lengths by the length of the secret.

[2] It is important to not confuse the randomness of the choice of h with the internal randomness of the encoder; the randomness of h comes from the *code construction*, and once a good choice of h is fixed once and for all, the encoder and decoder are properly defined and provide the expected guarantees.

function, and thus we finally obtain a fully explicit construction of efficient secret sharing schemes with nearly optimal parameters in all aspects. Namely, our main result in this work can be stated as follows.

Theorem 1 (Corollary 14, 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.* □

Another feature of our work is its complete modularity and simplicity which can help retain the practicality of Shamir's scheme. Our main result (Theorem 8) can be applied to any linear secret sharing scheme based on linear error-correcting codes that provides privacy via a dual distance argument (Shamir's original scheme being a special case). As a result, we are able to instantiate the result with virtually any algebraic family of linear list decodable codes, and particularly do so for the cases of (folded) Reed-Solomon and algebraic geometry codes. In contrast, the result of Cramer et al. [6] is only presented and proven when the underlying code is an algebraic geometry code, as the main goal of [6] is to obtain constant share lengths. Furthermore, as discussed above, [6] only provides a Monte-Carlo construction, since the choice of the hash function that pre-processes the secret is random which, in addition to adding to the description complexity of the final scheme, may cause the entire scheme to fail with an unfortunate choice of the (unverifiable) random hash function.

Same as Shamir's scheme and [24], 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 [25] (even against local, or oblivious, adversaries [2]), 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 [18]). Using algebraic geometry codes, we can prove the following.

Theorem 2 (Corollary 18, 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.* □

Previously, the best known construction achieving small share length was due to Cramer et al. [6] in which the share length is $\Theta(1 + \log(1/\eta)/n)$ and thus grows with the security parameter (see Table 1). Moreover, as mentioned above, this construction is not fully explicit and requires a randomly chosen hash function that is fixed once and for all and there is no clear efficient way of explicitly finding an appropriate hash function.

The efficiency of the scheme in Theorem 2 is dictated by the efficiency of the underlying list decoding algorithm for algebraic geometry codes. The encoding and list decoding algorithms in [18] 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.

We remark that the natural idea of reducing share length by using algebraic geometry codes rather than Reed-Solomon codes in secret sharing schemes dates back to a result of Chen and Cramer [5] and has been extensively studied since (cf. [9]), especially in the context of arithmetic secure multiparty computation.

It should be pointed out that, as discussed before, the focus of the present work is in showing that a simple modification of the existing Shamir's secret sharing scheme (i.e., the idea of amending the secret with an AMD tag that was actually proposed in [8] and shown to provide robustness in the sense of error-detection) essentially makes it optimally robust. This means that *existing systems* employing Shamir's scheme can be easily modified to provide stronger robustness against tampering adversaries, and this can be a very appealing improvement for practitioners that use Shamir's or related coding-theoretic schemes. In contrast, graph-based constructions such as [3,10,21] pursue a very different approach. Another appealing feature of coding-theoretic constructions such as our construction and Shamir's original scheme is that they allow an imbalance between the adversarial leakages and corruptions. In particular, for our constructions the adversary can read any τ fraction of the shares and use this information to corrupt any δ (whether including the shares previously read or not) fraction of the shares, and both privacy and robustness can be guaranteed as long as $\tau + \delta$ is nontrivially bounded away from 1.

Organization. The rest of the article is organized as follows. We explain the notation in Sect. 1.3. Preliminaries, including the exact notion of secret sharing schemes that we use in this work, are discussed in Sect. 2. Our general construction is presented and analyzed in Sect. 3. We then instantiate the construction using folded Reed-Solomon codes in Sect. 4.1 and folded algebraic geometry codes in Sect. 4.2. Finally, Sect. 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 Enc, 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[3] to $\mathcal{D}$. The scheme satisfies perfect privacy if $\epsilon = 0$.
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 $s \in \mathbb{F}_2^k$

[3] 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., $\epsilon = 0$.

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*. □

Secret sharing schemes that do not have a sharp threshold are known in the literature as *ramp* schemes, and the parameter $t + g + 1$ is sometimes called *reconstructability* parameter (cf. [6]).

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*). [8] A binary coding scheme (Enc, 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 [8], which we shall use in our constructions. Although, as stated in [8], 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 [8, Corollary 1] *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 $f : \mathbb{F}_2^k \times \mathbb{F}_2^q \to \mathbb{F}_2^q$ such that $q = \log(1/\eta) + \log(3 + k/\log(1/\eta)) + 1 = O(\log(k/\eta))$.

We note that explicit constructions of AMD codes are known that are better than the above for certain ranges of the parameters (e.g., [13]). However, the tag length of the construction in Theorem 6 is optimal within a constant factor of two, which suffices for our purposes. Furthermore, this construction is essentially based on Reed-Solomon codes (namely, the tag simply consists of a random point and evaluation of a polynomial defined by the message at that point), which fits nicely for use alongside a Shamir-type secret sharing scheme.

The notion of folded codes, following a line of work in algebraic list decoding (originally defined in [15]) 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

$$(c_1, \ldots, c_n) \in \mathcal{C}' \text{ if and only if } ((c_1(1), \ldots, c_1(m)), \ldots, (c_n(1), \ldots, c_n(m)) \in \mathcal{C},$$

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*

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

Then, there is an efficient and perfectly private robust secret sharing scheme (Share, 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

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

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 $f : \mathbb{F}_2^{n_0} \to \mathcal{C}''$ 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 $f' : \mathbb{F}_Q^n \to \mathbb{F}_2^{n_0}$ 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 (Share, 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 $Y = Y'$ (i.e., invoke error detection) 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 [22, Sect. 4.1], and since $\dim \mathcal{C}_0' < \dim \mathcal{C}_0$, we have $tm + 1 \leq nm - \dim \mathcal{C}_0'^{\perp} + 1 = \dim \mathcal{C}_0' + 1 = R'nm + 1 \leq Rnm - m + 1$. 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 [16] 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. □

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.

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 11 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 $(\alpha_{i,j} : i \in [n], j \in [m])$ 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 [15], folded Reed-Solomon codes attain an optimal trade-off between rate and list decoding radius. Specifically, the following is the main result proven[4] in [15].

[4] As stated in [15], 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.

Theorem 12 [15, follows from[5] Theorem 4.4] *For any constant parameter $\rho \in (0,1)$, $c \geq 1$, and integers $n > k > 0$, there is a $p_0 = O(nc/\rho^2)$ such that for any prime power $p \geq p_0$, there is an $\mathbb{F}_p$-linear folded Reed-Solomon code with message length k and block length n such that for some $\delta \geq 1 - k/n - \rho$, the following hold: (1) The code is list decodable from any δ fraction of errors with list size at most L, for some $L = p^{\Theta(\log(1/\rho)/\rho)}$; (2) The alphabet size of the code is $L^{c/\rho}$; (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 13 *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 $k_0 = O(\frac{\log(1/\rho)}{\nu\rho} \log(\frac{n}{\nu\rho}))$ such that for any integer $k \geq k_0$ the following holds. There is an efficient and perfectly private secret sharing scheme (Share, Rec) *with n shares, secret length k, privacy parameter t, threshold property with gap g, and share length q satisfying $(1+g-\nu)q \leq k + O(\log(k/\eta))$. Moreover, the scheme achieves a robustness parameter of δn and robustness error η.*

Proof Let c_0 be the constant from Theorem 8 and define $c := \lceil 2c_0\rho/\nu \rceil$. Let $\mathcal{C} \subseteq \mathbb{F}_Q^n$ be an $\mathbb{F}_p$-linear folded Reed-Solomon code, where $p = \Omega(nc/\rho^2)$ is a power of two to be determined later, as obtained by Theorem 12, of length n, message length $k' := t+g+1$, rate $R := k'/n$, and alphabet size that is list decodable from any $1 - R - \rho = \delta$ fraction of errors with list size bounded by $L = p^{\Theta(\log(1/\rho)/\rho)}$. Moreover, we set the alphabet size of the code is to be $Q = L^{c/\rho}$.

We instantiate Theorem 8 with the code $\mathcal{C}$ to obtain a secret sharing scheme (Share, Rec) with share length $q = \log Q = c \log L/\rho$. We now verify that the requirements of Theorem 8 are satisfied for any suitable choice of the secret length k.

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 $\mathsf{Enc}_{\mathcal{C}} : \mathbb{F}_Q^{k'} \to \mathbb{F}_Q^n$ 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 11). 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 $\mathbb{F}_{Q^{1/m}}$) 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 have

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

[5] The construction and analysis for this result is precisely as in [15] with the following additional considerations: The construction of [15] considers an m-level folding of the Reed-Solomon code with the smallest possible unfolded alphabet size (which is nm) and $m = O(1/\rho^2)$. Here, we consider an additional parameter $c \geq 1$ and allow a larger folding of $m = O(c/\rho^2)$, to have additional control over the alphabet size of the folded code compared with its list size. We furthermore allow the unfolded alphabet size p to be possibly larger than the minimum required size of nm. Finally, we upper bound the term $1/R$ in [15] by $O(1/\rho)$, noticing that when R is small, it is always possible to design a code at a slightly higher rate first and then truncate the message space to the desired length k.

By (1) in the statement of Theorem 8, and the above result, we wish to choose the share length q so as to satisfy the requirement

$$(g+1)q \geq k + c_0 \log(kL/\eta), \tag{4}$$

which can be rewritten, using the expression for q, as

$$(g+1)(c/\rho - c_0)\log L \geq k + c_0 \log(k/\eta).$$

Since $c/\rho \geq 2c_0 = \Omega(1)$ and the bound on the list size L satisfies $\log L = \Theta(\log(1/\rho)\log p/\rho)$, we may pick a large enough p so that the above inequality is satisfied, given the secret length k and the parameter η. In particular, we will choose p to be the smallest power of two that satisfies the above. From (4), we see that the share length q is upper bounded as follows

$$(g+1)q = k + O(\log(k/\eta)) + c_0 \log L.$$

Now we recall that, from the choice of c,

$$c_0 \log L = \rho c_0 q/c < \nu q,$$

and thus

$$(g+1-\nu)q = k + O(\log(k/\eta)),$$

as desired. We note that the minimum possible alphabet size Q is, according to Theorem 12, $(nc/\rho^2)^{\Theta(c\log(1/\rho)/\rho^2)}$. The logarithm of this quantity determines the minimum possible share length, and consequently the minimum allowed secret length k_0 in the statement of the theorem. Finally, to verify the threshold property, by Theorem 8 we have that the gap achieved by the code is upper bounded by $n - t - d = n - t - (n - k' + 1) = g$. This concludes the proof. □

We remark that for any (not necessarily robust) secret sharing scheme with threshold property and gap g, it is known that the share length q must satisfy $q \geq k/(1+g)$ (cf. [6]). Therefore, the share length achieved by Theorem 13 is essentially optimal.

For the important special case of $\delta = t/n$ and $g = 0$ we derive the following immediate corollary from Theorem 13.

Corollary 14 *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 a $k_0 = O_\nu(\log n)$ such that for any integer $k \geq k_0$, there is an efficient and perfectly private secret sharing scheme* (Share, Rec) *with n shares, secret length k and share length q, where $q(1-\nu) \leq k + O(\log(k/\eta))$. 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 14 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 [18]. 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 15 [18, 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.*

We now instantiate the general construction of Theorem 8 using the above result.

Theorem 16 *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*

$$\frac{k + c_0 \log k}{nq} + \frac{t}{n} + \delta \leq 1 - \rho - c_0 \frac{\log(1/\eta)}{nq} \tag{5}$$

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 13, but uses the folded algebraic geometry codes of Theorem 15 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[6] 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 15 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 ([26, 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 [26, 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

[6] Even though Theorem 15 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.

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

it suffices to ensure that

$$\frac{k}{nq} + \frac{t}{n} + \delta + \frac{c_0 \log(k/\eta)}{nq} \leq 1 - 3\rho' - \frac{c_0 \log L}{nq} - \frac{5}{n}. \tag{6}$$

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 (6) at least $1 - \rho$. Consequently, assuming (5), i.e.,

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

we have (6) 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 17 *Let c_0 be the constant from Theorem 8. 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*

$$\frac{k + c_0 \log k}{nq} + \gamma + \delta \leq 1 - \rho, \tag{7}$$

There is a perfectly private secret sharing scheme (Share, 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 $\eta = \exp(-\Omega(\rho nq))$. The scheme is efficient given polynomial (in n) amount of pre-processed information about the scheme.*

Proof We simply apply Theorem 16 with constant $\rho' := \rho/2$ (for the parameter ρ required by Theorem 16) to obtain a secret sharing scheme (Share, 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 $\eta = \exp(-\Omega(\rho nq))$ so that $c_0 \log(1/\eta) \leq \rho nq/2 - 1$, and thus

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

as needed by Theorem 16. 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 17, in turn, immediately implies the following result on robust secret sharing with privacy and robustness parameter δn for any $\delta < 1/2$.

Corollary 18 *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* (Share, 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 $\eta = \exp(-\Omega(\rho nq))$*, 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 14 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[7] so that $n \geq k/(\rho q)$. In Sect. 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 $n/2$.

Limitations of the method. As we have shown in this section, our framework can lead to robust secret sharing schemes against any fixed τ fraction of leaked shares (privacy) and any fixed δ fraction of corruptions (robustness) as long as $\tau + \delta < 1$, and we obtain a nearly optimal guarantee in terms of the share length in all cases. It is natural to ask whether $\rho := 1 - \delta - \tau$ can be made sub-constant. For example, in the maximum corruption scenario, where t shares are observed and corrupted for $n = 2t + 1$, we have $\rho = 1/n$, and more generally for $n = 2t + c$ we have $\rho = c/n$. To provide such guarantees, MDS-type list decodable codes of rate R and robust against any $1 - R - \rho$ fraction of errors with small list sizes will be required. Currently, the state of the art in explicit constructions of linear list decodable codes does not obtain sharp guarantees in the sub-constant ρ regime. Furthermore, general combinatorial negative bounds are known for any (even nonlinear) list decodable code. A simple probabilistic argument shows that, for any ρ, there are list decodable codes that achieve a list size of at most $1/\rho$ and alphabet size $\exp(O(1/\rho))$ [14]. Furthermore, an alphabet size of $\exp(\Omega(1/\rho))$ is necessary even for nonlinear codes [17, Chap. 3].

Currently, it is not known whether there are linear MDS-type codes matching the list decoding guarantees of fully random codes for the range of parameters discussed above. However, even if this turns out to be the case, the above-mentioned combinatorial lower bound on the alphabet size limits the allowed share lengths for the resulting secret sharing scheme. For Shamir's original scheme (as well as our scheme based on Reed-Solomon codes), the number of shares n for share length k can be at most $\exp(O(k))$, which is a reasonable restriction for cryptographic purposes (in other words, the minimum allowed share length for a give number of shares n while preserving the zero overhead in the share length is $\Omega(\log n)$). If we instantiate our result with an optimal list decodable code achieving $\rho = c/n$, the share length becomes $\Theta(1/\rho) = \Theta(n/c)$, which means that the minimum allowed secret length k becomes $\Omega(n/c)$. In other words, for constant c (i.e., the maximum robustness regime of $n = 2t + c$) our share length must be $\Omega(n)$, whereas for the non-robust Shamir's scheme, the share length (which is equal to the secret length) can be as small as $\log n$. Note that, for this regime, the linear dependence of the share length on n is not due to the overhead being sub-optimal (the overhead always remains nearly optimal since it depends on the list size and not the alphabet size of the code). The dependence is simply due to the restriction on the allowed secret length k (which must be at least k_0 for some $k_0 = \Omega(n/c)$).

[7] Note such a requirement is not a barrier for the Reed-Solomon based constructions such as Shamir's scheme and the result of Theorem 13, since we have $q \geq k$ in those schemes.

4.3 Optimality

In this section we briefly demonstrate that, for a general share length q, a robust secret sharing scheme satisfying (7) 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 [11,27], 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. [4]) of the form

$$I(S; Y') \leq \epsilon k,$$

where Y' is the wiretap channel's output, for arbitrarily small $\epsilon > 0$. A much stronger notion is *semantic security* (formalized in [4]) 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 [11,20], 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 (7) is the best to hope for.

Acknowledgements The author thanks Ronald Cramer, Venkatesan Guruswami, Rei Safavi-Naini, and Daniel Wichs, for illuminating discussions on the their related work.

References

1. Blakley G.R.: Safeguarding cryptographic keys. In: National Computer Conference, vol. 48, pp. 313–317. Springer (1979).
2. Bishop A., Pastro V.: Robust secret sharing schemes against local adversaries. In: Proceedings of Public-Key Cryptography (PKC), pp. 327–356 (2016).
3. Bishop A., Pastro V., Rajaraman R., Wichs D.: Essentially optimal robust secret sharing with maximal corruptions. In: Proceedings of the 35th Annual International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT 2016), pp. 58–86 (2016).
4. Bellare M., Tessaro S., Vardy A.: Semantic security for the wiretap channel. In: Proceedings of Advances in Cryptology CRYPTO 2012, Lecture Notes in Computer Science, vol. 7417, pp. 294–311. Springer (2012).
5. Chen H., Cramer R.: Algebraic geometric secret sharing schemes and secure multi-party computations over small fields. In: Advances in Cryptology—CRYPTO 2006, Lecture Notes in Computer Science, vol. 4117, pp. 521–536. Springer (2006).
6. Cramer R., Damgård I., Döttling N., Fehr S., Spini G.: Linear secret sharing schemes from error correcting codes and universal hash functions. In: Advances in Cryptology—EUROCRYPT 2015, Lecture Notes in Computer Science, vol. 9057, pp. 313–336. Springer (2015).
7. Cramer R., Damgård I., Fehr S.: On the cost of reconstructing a secret, or VSS with optimal reconstruction phase. In: Proceedings of Advances in Cryptology CRYPTO 2001, Lecture Notes in Computer Science, vol. 2139, pp. 503–523. Springer (2001).
8. Cramer R., Dodis Y., Fehr S., Padró C., Wichs D.: Detection of algebraic manipulation with applications to robust secret sharing and fuzzy extractors. In: Advances in Cryptology - EUROCRYPT 2008, Lecture Notes in Computer Science, vol. 4965, pp. 471–488. Springer (2008).
9. Cramer R., Damgård I., Nielsen J.B.: Secure Multiparty Computation and Secret Sharing. Cambridge University Press, Cambridge (2015).
10. Cevallos A., Fehr S., Ostrovsky R., Rabani Y.: Unconditionally-secure robust secret sharing with compact shares. In: Proceedings of Advances in Cryptology EUROCRYPT 2012, Lecture Notes in Computer Science, vol. 7237, pp. 195–208. Springer (2012).
11. Csiszár I., Körner J.: Broadcast channels with confidential messages. IEEE Trans. Inf. Theory **24**(3), 339–348 (1978).
12. Cabello S., Padró C., Sáez G.: Secret sharing schemes with detection of cheaters for a general access structure. In: Proceedings of Fundamentals of Computation Theory, Lecture Notes in Computer Science, vol. 1684, pp. 185–194. Springer (1999).
13. Cramer R., Padró C., Xing C.: Optimal Algebraic Manipulation Detection Codes in the Constant-Error Model, pp. 481–501. Springer, Berlin (2015).
14. Elias P.: Error-correcting codes for list decoding. IEEE Trans. Inf. Theory **37**(1), 5–12 (1991).
15. Guruswami V., Rudra A.: Explicit codes achieving list decoding capacity: error-correction with optimal redundancy. IEEE Trans. Inf. Theory **54**(1), 135–150 (2008).

16. Guruswami V., Smith A.: 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), pp. 723–732 (2010).
17. Guruswami V.: Algorithmic results in list decoding. Found. Trends Theor. Comput. Sci. **2**(2), 107–195 (2007).
18. Guruswami, V., Xing, C.: Optimal rate list decoding of folded algebraic-geometric codes over constant-sized alphabets. In: SODA, pp. 1858–1866 (2014).
19. Ishai Y., Ostrovsky R., Seyalioglu H.: Identifying cheaters without an honest majority. In: Proceedings of Theory of Cryptography (TCC 2012), Lecture Notes in Computer Science, vol. 7194, pp. 21–38. Springer (2012).
20. Leung-Yan-Cheong S.: On a special class of wiretap channels (corresp.). IEEE Trans. Inf. Theory **23**(5), 625–627 (1977).
21. Rabin T., Ben-Or M.: Verifiable secret sharing and multiparty protocols with honest majority. In: Proceedings of the Twenty-first Annual ACM Symposium on Theory of Computing (STOC '89), pp. 73–85 (1989).
22. Roth R.M.: Introduction to Coding Theory. Cambridge University Press, Cambridge (2006).
23. Shamir A.: How to share a secret. Commun. ACM **22**(11), 612–613 (1979).
24. Safavi-Naini R., Wang P.: A model for adversarial wiretap channels and its applications. J. Inf. Process. **23**(5), 554–561 (2015).
25. Stinson D.R.: An explication of secret sharing schemes. Des. Codes Cryptogr. **2**(4), 357–390 (1992).
26. Stichtenoth H.: Algebraic Function Fields and Codes, 2nd edn. Springer, Berlin (2009).
27. Wyner A.D.: The wire-tap channel. Bell Syst. Tech. J. **54**, 1355–1387 (1975).