A Framework for the Design of Secure and Efficient Proofs of Retrievability

Françoise Levy-Dit-Vehel, Maxime Roméas

A Framework for the Design of Secure and Efficient Proofs of Retrievability

Françoise Levy-dit-Vehel[1] and Maxime Roméas[2]

[1] LIX, ENSTA Paris, INRIA, Institut Polytechnique de Paris, 91120 Palaiseau, France. levy@ensta.fr
[2] LIX, École polytechnique, INRIA, Institut Polytechnique de Paris, 91120 Palaiseau, France. romeas@lix.polytechnique.fr

Abstract. Proofs of Retrievability (PoR) protocols ensure that a client can fully retrieve a large outsourced file from an untrusted server. Good PoRs should have low communication complexity, small storage overhead and clear security guarantees with tight security bounds. The focus of this work is to design good PoR schemes with simple security proofs. To this end, we propose a framework for the design of secure and efficient PoR schemes that is based on Locally Correctable Codes, and whose security is phrased in the Constructive Cryptography model by Maurer. We give a first instantiation of our framework using the high rate lifted codes introduced by Guo *et al.* This yields an infinite family of good PoRs. We assert their security by solving a finite geometry problem, giving an explicit formula for the probability of an adversary to fool the client. Moreover, we show that the security of a PoR of Lavauzelle and Levy-dit-Vehel was overestimated and propose new secure parameters for it. Finally, using the local correctability properties of Tanner codes, we get another instantiation of our framework and derive an analogous formula for the success probability of the audit.

1 Introduction

1.1 Context and state-of-the-art

With the continuous increase in data creation, individuals and business entities call upon remote storage providers to outsource their data. This new dependency raises some issues, as the storage provider can try to read and/or modify the client's data. Besides, when a client does not often access his data, the service provider can delete it to make room for another client's data. In this context, it appears important to deploy client side protections designed to bring security guarantees like confidentiality and integrity. In this work, we focus on the following problem: given a client who stored a file on a server and erased its local copy, how can he check if he is able to retrieve his file from the server in full? Addressing this issue is the goal of a class of cryptographic protocols called Proofs of Retrievability (PoRs).

The first PoR scheme was proposed in 2007 by Juels and Kaliski [8] and was based on checking the integrity of some sentinel symbols secretly placed by

the client before uploading its file. This scheme has low communication but its drawback is that it is bounded-use only, as the number of possible verifications depends on the number of sentinels. To correct this drawback, Shacham and Waters [15] proposed to append some authenticator symbols to the file. Verification consists in checking random linear combinations of file symbols and authenticators. Then comes a few PoR schemes based on codes. Bowers *et al.* [2] proposed a double-layer encoding with the use of an inner code to recover information symbols and an outer code to correct the remaining erasures. Dodis *et al.* [4] formalize the verification process as a request to a code which models the space of possible answers to a challenge. They use Reed-Solomon codes to design their PoR scheme. In 2013, Paterson *et al.* [14] laid the foundation for studying PoR schemes using a coding theoretic framework. Following these ideas, Lavauzelle and Levy-dit-Vehel [11] (2016) used the local structure of the lifted codes introduced by Guo *et al.* [5] to build a PoR scheme, that compares favourably to those presented above w.r.t. storage overhead.

Unfortunately, PoR schemes have a few issues. Indeed, their security definitions are often unclear, making it hard to understand what they really achieve. Moreover, when a client wants to retrieve his data, the security guarantees brought by the use of the PoR scheme only holds under the condition that both client and server unveil some private information (client's secret material and server's state). We give a detailed explanation of this in sec. 2.2. In 2018, Badertscher and Maurer [1] used the Constructive Cryptography (CC) model [12] to propose a new PoR definition, that avoids the aforementioned flaws. They also designed a PoR scheme based on generic erasure codes. Generalizing [1] and [11], we introduce a framework for designing secure, composable and efficient PoR protocols based on locally correctable codes.

Our approach allows us to design and study the security of PoR schemes in a modular fashion, that achieves stronger security and clearer security guarantees than previous schemes (whose security was based on so-called ϵ adversaries or related notions). Using another definitional model such as the Universal Composability one by Canetti [3] would probably give closely related results. We chose to use CC because it makes the resources available to the parties (namely, untrusted server storage, local memories, communication channels) explicit. It also makes the switching between computational, statistical and information-theoretic security notions easy.

Locally correctable codes (LCCs), which are at the core of our work, were formally introduced by Katz and Trevisan [9] in 2000. Reed-Muller codes are well known to be locally correctable, but with poor rate as their length grows. The year 2011 has seen a breakthrough in the theory of codes with locality, with the construction by Kopparty *et al.* [10] of a class of high-rate LCCs -the multiplicity codes- generalizing the Reed-Muller class. Other high-rate LCCs are notably the lifted codes introduced by Guo *et al.* [5], and the expander codes of Hemenway *et al.* [6]. The high rates of these codes permit to minimize the server storage overhead, making them best suited for the outsourced storage context. We give an instantiation of our framework using the lifted codes of Guo *et al.*.

In a nutshell, we exploit the geometric properties of lifted codes and the CC security model for PoRs to give simple security proofs with tight bounds. This is a key difference between our approach and the one of Lavauzelle and Levy-dit-Vehel [11], which is also based on lifted codes and can in fact be seen as a different instantiation of our framework. We also propose another instantiation of our framework using the graph codes of Tanner [16].

1.2 Contributions

Given a LCC, we propose a canvas for deriving a PoR scheme. We get efficiency by taking advantage of the local correctability of the code to design an audit with low communication complexity. Using the CC security model for PoRs of [1], we give clear and composable security guarantees for our PoR construction. We are also able to give security bounds derived from geometric/combinatorial proofs and we reevaluate the security of the scheme of [11].

As in many protocols, the client first encodes its file and uploads it to the server. Retrieving the file is done by iterating the local decoder. With such a decoding process in mind, for extracting the file, we identify the adversarial configurations of corruptions that would prevent its extraction. This analysis of adversarial impact permits us to phrase the security of the audit -which heavily relies on the local correction step- as a problem about the structure of the code. Namely, if the code uses geometric properties of $\mathbb{F}_q^m$, we reduce the security of the audit to a finite geometry problem that we thoroughly address in sec. 5.2. If the code is a graph code, we reduce the security to a graph theoretic problem in sec. 5.3.

Instantiating our framework with the lifted Reed-Solomon codes: we characterize all the configurations of corruptions that are impossible to correct using the local correctability of those codes. More precisely, we show that these configurations of corruptions correspond to sets of points verifying a geometric property inside a vector space over a finite field. Then, we show that these sets of points belong to a large number of affine lines. From this we derive an explicit formula for the probability of the adversary to fool the client. Thus, we get a family of PoR schemes with precise security guarantees. Efficiency of this construction is shown in fig. 6 and 7 of sec. 6, where we also give a comparison between our parameters and those of the PoR scheme of [11]. Our security analysis shows that their scheme is insecure for their choice of parameters. Fortunately, we are able to give new ranges of secure parameters in fig. 6.

Instantiating our framework with Tanner codes: we proceed analogously as in the lifted Reed-Solomon codes case to first design a global decoder, and then to characterize the configurations of erased edges that are unrecoverable by the decoding algorithm. This way, we derive a bound for the failure probability of the audit, that only depends on the choice of the graph.

In order to design our framework, we constructed a protocol to authenticate outsourced data (**aSMR**, for Authentic Server Memory Resource[3]) that is tai-

[3] Following the terminology of [1].

lored for PoR purposes (see sec. 4). Our **aSMR** is different from the one of [1] in several aspects, notably, when dealing with encoded data, we halve the extra storage needed in comparison to [1]. Further details are to be found in sec. 4. This new construction might be useful in other code-based schemes. It can also be used to improve the efficiency of the generic PoR of [1].

2 Background

2.1 The Constructive Cryptography model

The CC model, introduced by Maurer [13] in 2011, aims at asserting the real security of cryptographic primitives. To do so, it redefines them in terms of so-called *resources* and *converters*. In this model, starting from a basic resource (e.g. communication channel, shared key, memory server...), a converter (a cryptographic protocol) aims at constructing an enhanced resource, *i.e.*, one with better security guarantees. The starting resource, lacking the desired security guarantees, is often called the *real* resource and the obtained one is often called the *ideal* resource, since it does not exist as is in the real world. An example of ideal resource is a confidential server, where the data stored by a client is readable by this client only. The only information that leaks to other parties is its length. This resource does not exist, but it can be emulated by an insecure server on which the client uses a suitable encryption protocol. We say that this *construction* of the confidential server is secure if the adversary in the ideal world, *i.e.* when a confidential server is used, is able to do the same things than in the real world setting, *i.e.* when an insecure server is used together with the protocol. We use the fact that the ideal world is by definition secure and contraposition to conclude. This construction notion is illustrated in fig. 2.

The CC model follows a top-down approach, allowing to get rid of superfluous hypotheses made in other models. A particularity of this model is its composability, in the sense that a protocol obtained by composition of a number of secure constructions is itself secure. We give the required material to understand how we use CC below. We follow the presentation of [7].

Resources, converters and distinguishers. A *resource* $\mathbf{R}$ is a system that interacts, in a black-box manner, at one or more of its *interfaces*, by receiving an input at a given interface and subsequently sending an output at the same interface. Do note that a resource only defines the observable behavior of a system and not how it is defined internally.

In CC, *converters* are used to link resources and reprogram interfaces, thus expressing the local computations of the parties involved. A converter is plugged in a set of interfaces at the inside and provides a set of interfaces at the outside. When it receives an input at its outside interface, the converter uses a bounded number of queries to the inside interface before computing a value and outputting it at its outside interface. A converter π connected to the interface set $\mathcal{I}$ of a resource $\mathbf{R}$ yields a new resource $\mathbf{R}' := \pi^{\mathcal{I}}\mathbf{R}$. The interfaces of $\mathbf{R}'$ inside the set $\mathcal{I}$ are the interfaces emulated by π. A protocol can be modeled as a tuple of converters with pairwise disjoint interface sets.

A *distinguisher* $\mathbf{D}$ is an environment that connects to all interfaces of a resource $\mathbf{R}$ and sends queries to them. At any point, the distinguisher can end its interaction by outputting a bit. Its advantage is defined as $\Delta^{\mathbf{D}}(\mathbf{R}, \mathbf{S}) := |\Pr[\mathbf{D}(\mathbf{R}) = 1] - \Pr[\mathbf{D}(\mathbf{S}) = 1]|$.

In this work, we make statements about resources with interface sets of the form $\mathcal{I} := \mathcal{P} \cup \{\mathsf{S}, \mathsf{W}\}$, where $\mathcal{P} := \{\mathsf{C}_0, \mathsf{C}\}$ is the set of honest client interfaces. A protocol is a pair of converters $\pi := (\pi_{\mathsf{C}_0}, \pi_{\mathsf{C}})$ that specifies one converter for each interface. The goal of this protocol is to construct a so-called ideal resource from an available real resource in presence of a potentially dishonest server S. The world interface W models the direct influence of a distinguisher on a resource.

Specifications. In CC, systems are grouped according to desired or assumed properties that are relevant to the user, while other properties are ignored on purpose. A *specification* $\mathcal{S}$ is a set of resources that have the same interface set and share some properties, for example confidentiality. In order to construct this set of confidential resources, one can use a specification of assumed resources $\mathcal{R}$ and a protocol π, and show that the specification $\pi\mathcal{R}$ satisfies confidentiality. Proving security is thus proving that $\pi\mathcal{R} \subseteq \mathcal{S}$, sometimes written as $\mathcal{R} \xrightarrow{\pi} \mathcal{S}$, and we say that the protocol π constructs the specification $\mathcal{S}$ from the specification $\mathcal{R}$. The composition property of the framework comes from the transitivity of inclusion. Formally, for specifications $\mathcal{R}, \mathcal{S}$ and $\mathcal{T}$ and protocols π for $\mathcal{R}$ and π' for $\mathcal{S}$, we have $\mathcal{R} \xrightarrow{\pi} \mathcal{S} \wedge \mathcal{S} \xrightarrow{\pi'} \mathcal{T} \Rightarrow \mathcal{R} \xrightarrow{\pi' \circ \pi} \mathcal{T}$.

We use the real-world/ideal-world paradigm, and often refer to $\pi\mathcal{R}$ and $\mathcal{S}$ as the real and ideal-world specifications respectively, to understand security statements. Those statements say that the real-world is "just as good" as the ideal one, meaning that it does not matter whether parties interact with an arbitrary element of $\pi\mathcal{R}$ or one of $\mathcal{S}$. This means that the guarantees of the ideal specification $\mathcal{S}$ also apply in the real world where an assumed resource is used together with the protocol.

We use *simulators*, *i.e.*, converters that translate behaviors of the real world to the ideal world, to make the achieved security guarantees obvious. For example, one can model confidential servers as a specification $\mathcal{S}$ that only leaks the data length, combined with some simulator σ, and show that $\pi\mathcal{R} \subseteq \sigma\mathcal{S}$. It is then clear that the adversary cannot learn anything more that the data length.

Server-memory resources. We recall the constructions of [1] that we will use or improve in this work. The first resource is the basic server-memory resource (SMR) denoted by $\mathbf{SMR}_{\Sigma,n}$ where Σ is the alphabet and n is the number of data blocks. It allows a client to read and write data blocks that are encoded as elements of a finite alphabet Σ via interface C. The interface C_0 is used to set up the initial state of the resource. The server can be "honest but curious" by obtaining the entire history of accesses made by the client (a log file) and reading its data at interface S_H. The server can also be intrusive and overwrite data using its interface S_I when the resource is set into a special write mode. This write mode can be toggled by the distinguisher at the world interface W. The specification of the resource $\mathbf{SMR}_{\Sigma,n}$ is given in fig. 1.

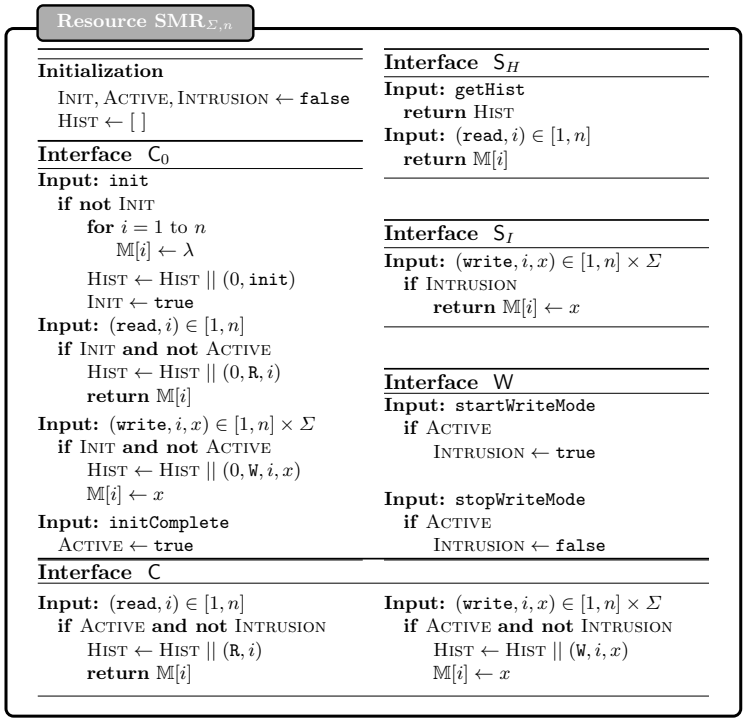

Fig. 1. Description of the basic server-memory resource

In fig. 2, taken from [1], we illustrate the CC construction notion on SMRs. The SMR security guarantees can be augmented to provide authenticity by using a suitable protocol in this construction notion. This new SMR is called authentic SMR, denoted by $\mathbf{aSMR}_{\Sigma,n}$, and is constructed in [1]. In **aSMR**, the behavior of the server at interface S_I is weakened as the server cannot modify the content of data blocks but is limited to either delete or restore previously deleted data blocks. A deleted data block is indicated by the special symbol ϵ. In this work, we use a different **aSMR** specification that the one used in [1]. We modify the `restore` behavior to only restore data blocks that were deleted after the last client update of the database. We introduce a version number that tracks the number of said updates in the history of the **aSMR** and the client is now allowed to overwrite corrupted data blocks. These changes decrease the storage overhead along with the communication complexity of read operations while the communication complexity of write operations is increased in comparison to the specification of [1]. Our changes to the **aSMR** yield substantial improvements for the parameters of our code-based PoR schemes. Our take on the **aSMR** resource is described in fig. 3 and our changes are precised in sec. 4.

2.2 Proofs of Retrievability

Proofs of Retrievability (PoR) are cryptographic protocols whose goal is to guarantee that a file stored by a client on a server remains retrievable in full. PoRs

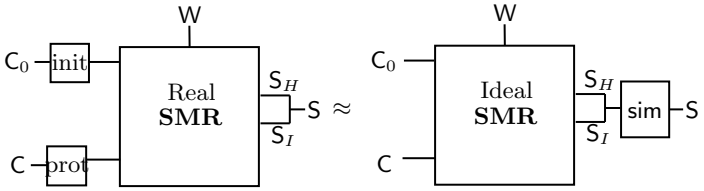

Fig. 2. Illustration of the construction notion for SMRs. On the left, we have a real **SMR** with a protocol for the client. On the right, we have an ideal **SMR** with stronger security guarantees. The construction is secure if there exists a simulator that makes these two resources indistinguishable.

Resource aSMR$_{\Sigma,n}$

The **aSMR** definition is identical to **SMR** except for the influence of an adversary at interface S_I and the addition of a version number ctr.

Interface C

Input: $(\texttt{read}, i) \in [1, n]$
if ACTIVE **and not** INTRUSION
HIST $\leftarrow$ HIST $||$ $(\texttt{R}, i)$
return $\mathbb{M}[i]$

Input: $(\texttt{write}, i, x) \in [1, n] \times \Sigma$
if ACTIVE **and not** INTRUSION
$ctr \leftarrow ctr + 1$
HIST $\leftarrow$ HIST $||$ $(\texttt{W}, i, x, ctr)$
$\mathbb{M}[i] \leftarrow x$

Interface S_I

Input: $(\texttt{delete}, i) \in [1, n]$
if INTRUSION
$\mathbb{M}[i] \leftarrow \epsilon$

Input: $(\texttt{restore}, i) \in [1, n]$
if INTRUSION
if $\exists k, x$:HIST$[k] = (\texttt{W}, i, x, ctr)$
$\mathbb{M}[i] \leftarrow x$

Fig. 3. Our new authentic SMR (only the differences with **SMR** are shown)

thus involve two parties: a client who owns a file F and a server, here modeled as a SMR, on which F is stored. We recall the commonly used definitions for PoR security as presented in [11]. A PoR scheme is composed of three main procedures:

- *An initialization phase.* The client encodes his file F with an initialization function $\mathbf{Init}(F) = (\tilde{F}, \texttt{data})$. He keeps $\texttt{data}$ (*e.g.* keys, etc.) for himself, then he sends $\tilde{F}$ to the server and erases F.
- *A verification phase.* The client produces a challenge c with a randomized **Chall** function and sends it to the server. The latter creates a response $r = \mathbf{Resp}(\tilde{F}, c)$ and sends it back to the client. The client checks if r is correct by running $\mathbf{Verif}(c, r)$, which also access $\texttt{data}$, and outputs $\texttt{accept}$ if r is considered correct and $\texttt{reject}$ otherwise.
- *An extraction phase.* If the client has been convinced by the verification phase, he can use his **Extract** algorithm to recover his whole file with high probability.

The security of PoR schemes is usually defined with ϵ-adversaries. In a PoR scheme, the client wants to use the **Verif** procedure to be sure that he will be able to retrieve his file in full when using the **Extract** procedure. The following definition models the fact that, if the server's answers to client's challenges make him "look like" he owns the file, then the client must be able to recover it entirely.

Definition 1 (ϵ-adversary). *Let $\mathcal{P}$ be a PoR system and X be the space of challenges generated by* **Chall**. *An ϵ-adversary $\mathcal{A}$ for $\mathcal{P}$ is an algorithm such that, for all files F,* $\Pr_{x \in X}[\mathbf{Verif}(x, \mathcal{A}(x)) = \texttt{false}] \leq \epsilon$

The client models the server as an ϵ-adversary and uses his verification process to maintain an approximation of ϵ. Depending on this estimate, the client can decide whether his file is retrievable or not. The security of PoRs is thus usually measured as follows:

Definition 2 (PoR security). *Let $\epsilon, \rho \in [0, 1]$. A PoR system is (ϵ, ρ)-sound if, for all ϵ-adversaries $\mathcal{A}$ and for all files F, we have* $\Pr[\mathbf{Extract}^{\mathcal{A}} = F] \geq \rho$ *where the probability is taken over the internal randomness of* $\mathbf{Extract}^{\mathcal{A}}$.

As pointed out by Badertscher and Maurer in [1], this model has a major drawback concerning client-side security guarantees. The most important thing for the client, the availability of his data, is conditioned to the execution of the **Extract** algorithm which needs to access the client's private data and the server's strategy (as indicated in def. 2). In practice, no server would reveal its entire state to a client. This problem is addressed in [1], where the authors used the CC framework to propose a definition of PoRs that fixes this drawback. In their work, they introduced an ideal abstraction of PoRs in the form of an ideal SMR that sees the client's interface augmented with an `audit` mechanism. On an `audit` request, the resource checks whether the current memory content is indeed the newest version that the client wrote to the storage. If a single data block has changed, the ideal audit will detect this and output `reject` to the client. In case of a successful audit (returning `accept`), this guarantee holds until the server gains write-access to the storage, in which case a new audit has to reveal whether modifications have been made. We present the specification of the auditable and authentic SMR $\mathbf{aSMR}^{\mathsf{audit}}_{\Sigma,n}$ in fig. 4. In addition to the advantages we discussed, we believe that this CC based security model is simpler and more intuitive than the one of ϵ-adversaries.

In CC, a PoR scheme is given by a pair of converters $\mathsf{por} := (\mathsf{por}_{\mathsf{init}}, \mathsf{por}_{\mathsf{audit}})$ where $\mathsf{por}_{\mathsf{init}}$ implements the $(\texttt{write}, F)$ query that uploads F (or an encoded/encrypted version of F) on the SMR, where F is the client's file and $\mathsf{por}_{\mathsf{audit}}$ implements the `audit` query that returns either `accept` or `reject`, and the `read` query that extracts the file F from the SMR.

2.3 Locally Correctable Codes

In [1], Badertscher and Maurer give a protocol based on generic erasure codes to construct the auditable **aSMR**. Due to the use of classical codes, a client who wants to read a single data block needs to read the entire memory in order for him to run the decoding algorithm of the code to recover (or not) the data block. In this work, we show how one can use LCCs, so that one has to read only a small number of memory positions to recover one data block, while keeping the auditable property of the constructed resource. We now briefly present LCCs, which were formally introduced by Katz and Trevisan [9] in 2000.

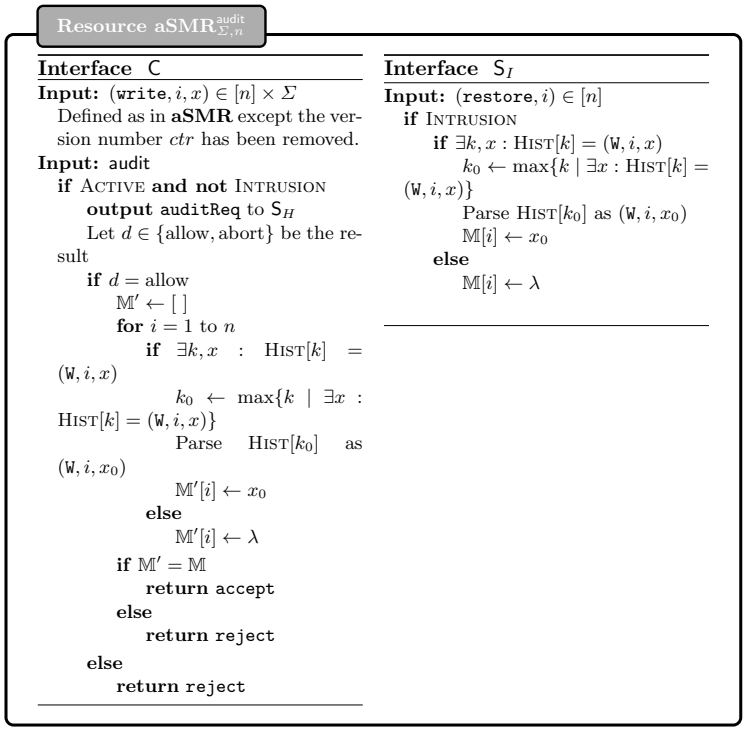

Fig. 4. Description of the auditable and authentic SMR of [1] (only the differences with our **aSMR** are shown)

Definition 3 (Locally correctable code). *Let $r \in \mathbb{N}, \delta \in [0,1]$ and $\epsilon : [0,1] \to [0,1]$. A code $\mathcal{C} \subseteq \mathbb{F}_q^n$ is said to be (r, δ, ϵ)-locally correctable if there exists a probabilistic decoding algorithm $\mathcal{A}$ such that,*

1. *For all $\mathbf{c} \in \mathcal{C}$, for all $i \in [\![1, n]\!]$ and for all vectors $\mathbf{y} \in \mathbb{F}_q^n$ with relative Hamming distance $\Delta(\mathbf{c}, \mathbf{y}) \leq \delta$, we have $\Pr[\mathcal{A}^{\mathbf{y}}(i) = \mathbf{c}_i] \geq 1 - \epsilon(\delta)$, where the probability is taken over the internal randomness of $\mathcal{A}^{\mathbf{y}}$.*
2. *The algorithm $\mathcal{A}$ makes at most r queries to the vector $\mathbf{y}$.*

In this work, we consider locally correctable codes for erasures and we do not use the estimate of their failure probability. See sec. 3 for more details.

3 Our framework

We describe our framework which derives PoR schemes from a given LCC $\mathcal{C}$. In all our PoRs, the client's file is encoded as a codeword of $\mathcal{C}$ and uploaded to the server. We want to protect the client from an adversary able to introduce corruptions on the outsourced file. To do so, we need to describe an audit that probes a few symbols of the outsourced file and accepts if it thinks that the corruptions can all be corrected. Recall that, in the CC definition, an audit is considered secure if it only succeeds when the outsourced file is retrievable in

full, without modifications. If we want to derive PoR schemes from an LCC $\mathcal{C}$ in CC, we thus need to do the following three things:

1. Give an extraction procedure that aims at retrieving the outsourced file while correcting any corruption encountered.
2. Characterize the configurations of corruptions that are uncorrectable by this extraction procedure.
3. Give an audit procedure that is able to detect those configurations of uncorrectable corruptions on the outsourced file.

Since a good PoR scheme must have low communication complexity, we want to exploit the locality of LCCs to design our audit procedure. We choose our extraction procedure as an iteration of the local correction algorithm of the LCC. This means that our schemes will try to locally correct any corruption encountered during the extraction. Thus, we need a way to identify those corruptions. Using the composability of the CC framework, we will place ourselves in a setting where adversaries can only introduce erasures on the outsourced file. We can design our PoR schemes with this assumption and we will need to construct an authenticated server to realize it later on. Our blueprint becomes:

1. Give an extraction procedure that aims at correcting erasures by using the local correctability of $\mathcal{C}$.
2. Characterize the configurations of erasures that are uncorrectable by this extraction procedure.
3. Our audit is the following: try to locally correct a random position of the outsourced file, if the correction is impossible return `reject`, else return `accept`.

In step 2, we identify the configurations of erasures that are unrecoverable when iterating the local correction of $\mathcal{C}$. We find a lower bound on the number of local correction queries that would fail if such a configuration of erasures existed. When instantiating our framework in sec. 5, we shall see that this problem is, in practice, much easier than giving a lower bound on the minimum size of such a configuration of unrecoverable erasures. In the CC model of security for PoRs, the advantage of the adversary in breaking the security of the scheme is the probability that the audit accepts while the file is not retrievable. In our case, our audit consists in checking if a random local correction query succeeds. Our file is not retrievable if there exists a configuration of unrecoverable erasures. Thus, the lower bound we computed above is all we need to assess the security of the PoR. We give a complete proof when instantiating our framework, see th. 1 of sec. 5.

More precisely, let $\mathcal{C}$ be an erasure code of length n, alphabet Σ and erasure symbol $\perp$. Suppose that $\mathcal{C}$ possesses a local erasure decoder $\mathcal{L}$ with query space $\mathcal{Q} \subseteq 2^{[1,n]}$. On query $q \in \mathcal{Q}$ and input $w \in (\Sigma \cup \{\perp\})^n$ such that there exists $c \in \mathcal{C}$ such that for any $i \in [1,n]$, $w_i \neq \perp$ implies $w_i = c_i$, $\mathcal{L}$ probes the symbols $w_{|q} := (w_i)_{i \in q}$ of w and attempts to correct its erasures if they exist. We can define a global decoder $\mathcal{G}$ for $\mathcal{C}$ by iterating $\mathcal{L}$ until no erasures remain. Let

P be a predicate on $\cup_{i=0}^{n}(\Sigma \cup \{\perp\})^i$, *i.e.*, for $w \in (\Sigma \cup \{\perp\})^n$ and $q \in \mathcal{Q}$, $P(w_{|_q}) \in \{\mathsf{true}, \mathsf{false}\}$. Let $0 \leq \epsilon \leq 1$ and suppose that we have the following property:

$$\forall w \in (\Sigma \cup \{\perp\})^n, \text{if at least one erasure of } w \text{ cannot be corrected by } \mathcal{G} \text{ then } \Pr_{q \in \mathcal{Q}}[P(w_{|_q}) = \mathsf{false}] \geq 1 - \epsilon \quad (1)$$

We define our general PoR scheme $\mathsf{por} := (\mathsf{por}_{\mathsf{init}}, \mathsf{por}_{\mathsf{audit}})$, where:

1. On input $(\mathtt{write}, F)$, $\mathsf{por}_{\mathsf{init}}$ encodes F into a codeword $\tilde{F}$ of $\mathcal{C}$ and writes $\tilde{F}$ in the **aSMR** memory.
2. On input $\mathtt{audit}$, $\mathsf{por}_{\mathsf{audit}}$ samples a query $q \in \mathcal{Q}$ uniformly. If w is the file stored in the SMR, $\mathsf{por}_{\mathsf{audit}}$ retrieves $w_{|_q}$ with $\mathtt{read}$ queries. Then, the converter returns $\mathtt{accept}$ if $P(w_{|_q} = \mathsf{true})$ and returns $\mathtt{reject}$ otherwise.
3. On input $(\mathtt{read})$, the converter $\mathsf{por}_{\mathsf{audit}}$ tries to extract the file F using the global decoder $\mathcal{G}$ of $\mathcal{C}$.

Recall that in the CC model of security for PoRs, the advantage of the adversary in breaking the security of por is the probability that the audit accepts while the file is not retrievable. In our case, this advantage is upper bounded by ϵ (see eq. 1). We believe our security model for PoRs to be cleaner, simpler and to give clearer security guarantees than the ϵ-adversary model.

4 Our authentication protocol

Recall that we focus on schemes based on erasure capabilities of error correcting codes. Thus, we need a setting where the actions of adversaries only lead to introducing erasures, instead of errors, in the outsourced data. This is exactly what an authentic server-memory resource (**aSMR**) achieves since the adversary can only delete data or restore previously deleted data. Thus, we need a protocol that constructs an **aSMR** from a basic **SMR**.

In [1], Badertscher and Maurer present a protocol that constructs an **aSMR** using a MAC function, timestamps and a tree structure on the outsourced data. Their construction of the **aSMR** has the following features:

1. The **aSMR** of size n with alphabet Σ is constructed from an **SMR** of size $2n - 1$, alphabet $\Sigma \times \mathbb{Z}_q \times \mathcal{T}$ and a local memory of constant size for the client. $\mathcal{T}$ is the tag space of the MAC function used.
2. To read or write one memory cell on **aSMR**, the protocol of [1] produces $O(\log n)$ $\mathtt{read}$ and $\mathtt{write}$ queries to **SMR**.

Our work focuses on PoR schemes where the client uploads a very large encoded file to an outsourced server. In this context, the logarithm of the size of the alphabet Σ is an order of magnitude smaller than the length of the MAC tags. The **aSMR** construction of [1] is thus not suited for this kind of application. Its issues are threefold. First, since the file size is huge, a factor of 2 in the storage

overhead is a big problem. Second, the $O(\log n)$ communication complexity for write operations is of no use to us since we will be working on encoded data and updating a codeword requires anyway to read a linear number of symbols. Third, the verification phase of PoRs often consists in probing as few symbols as possible to ensure that the outsourced file is retrievable in full. Having a $O(\log n)$ read communication complexity is a problem in this context.

With these observations, we now present a different protocol that constructs an **aSMR** with good features for our context:

1. Our **aSMR** of size n with alphabet Σ is constructed using an **SMR** of size n, alphabet $\Sigma \times \mathcal{T}$ and a local memory of constant size for the client.
2. A `read` request to our **aSMR** produces only one `read` request to **SMR**.
3. A `write` request to our **aSMR** produces at most $2n - 1$ `read` and `write` requests to **SMR**.

This way, we minimize the storage overhead and the communication complexity of read requests on the one hand. On the other hand, the increased communication complexity for write requests does not matter since our PoR schemes use only one such request. We sketch our protocol.

In the following, let n be the size of the **SMR**, $f_{sk}(\cdot)$ be a MAC function with tag space $\mathcal{T}$ and Σ be a finite alphabet. The protocol `auth` starts with the clients choosing a secret key sk for the MAC function, setting a version number ctr to 0. The main idea is the following: if the i-th cell is supposed to store the data $x \in \Sigma$, the protocol will store the pair $(x, f_{sk}(x, ctr, i)) \in \Sigma \times \mathcal{T}$ instead. Do note that the version number ctr is incremented with every write request. This also means that every valid tag needs to be updated with every write request. Intuitively, this protocol prevents the adversary from:

1. Replacing the data x with $y \neq x$ since this would make the tag invalid.
2. Moving the data stored in location i to location $j \neq i$ since this would make the tag invalid.
3. Replaying an older value since the version numbers would not match and the tag would thus be invalid.

5 Instantiation with high rate LCCs

5.1 Lifted Reed-Solomon Codes

We introduce a very interesting class of LCCs, namely the high rate lifted Reed-Solomon (RS) codes of Guo *et al.* [5]. In the following, let $\mathbb{F}_q$ be the finite field with q elements and m be a positive integer. The set of affine lines in $\mathbb{F}_q^m$ is denoted by $\mathcal{L}_m := \{(at + b)_{t \in \mathbb{F}_q} \mid a, b \in \mathbb{F}_q^m\}$. $\mathrm{RS}_q[q, d]$ is the q-ary RS code of length q and minimum distance $d = q - k + 1$.

Definition 4 (Lifted Reed-Solomon Code [5]). *Let $\mathbb{F}_q$ be a finite field. Let $d, m \in \mathbb{N}^*$. The m-lift of $\mathrm{RS}_q[q, d]$ is $\mathrm{Lift}_m(\mathrm{RS}_q[q, d]) := \{w \in \mathbb{F}_q^m \mid \forall$ line $\ell \subseteq \mathbb{F}_q^m, w_{|\ell} \in \mathrm{RS}_q[q, d]\}$.*

As we are using an **aSMR**, codewords can only be affected by potential erasures. A codeword of the RS base code $\mathrm{RS}_q[q,d]$ is the vector of evaluations of a polynomial f of degree strictly less than $k = q - d + 1$. Thus, if there are at most $d-1$ erasures, we can always recover the codeword *i.e.* the polynomial f by interpolating on $k > \deg f$ points. Therefore, if we want to correct a coordinate $x \in \mathbb{F}_q^m$ of the $\mathrm{Lift}_m(\mathrm{RS}_q[q,d])$ code, we can pick a random line going through x and run the aforementioned local decoding algorithm.

5.2 The lifted RS PoR scheme

In this section, we use our PoR framework to design a secure and efficient PoR scheme using lifted RS erasure codes. We call this scheme *lifted RS PoR scheme*. We build our PoR for an **aSMR** and then use the composability of CC. Since this server is authenticated, we only have to deal with potential erasures instead of errors. Using the blueprint of sec. 3, we need to do the following:

1. Give a global decoding algorithm for lifted RS codes using their local correctability.
2. Characterize the configurations of erasures that are unrecoverable by this algorithm.
3. Give an audit procedure that is able to detect those configurations of uncorrectable corruptions on the outsourced file.

Let us start with the global decoding algorithm. For the lifted RS code $\mathrm{Lift}_m(\mathrm{RS}_q[q,m])$, our global decoder works as follows. For each erasure, the decoding algorithm corrects it by finding, if it exists, a line going through the erasure and containing less than $d-1$ other erasures (using interpolation as quoted in sec. 5.1). If one or more erasures have been corrected during this step, the algorithm tries to correct the remaining erasures using the same method. Indeed, since some erasures were corrected, there exist lines with less erasures than before. If, during one iteration, no erasures have been corrected, the algorithm stops and returns the current vector. We give a pseudo-code description of this algorithm in fig. 5.

Input: The encoded file V with potential erasures
Output: The encoded file $\tilde{F}$.
  repeat
    $E := \emptyset$
    for an erased position $x \in \mathbb{F}_q^m$
      if there exists a line $\ell \subseteq \mathbb{F}_q^m$ going through x with strictly less than d erasures.
        Use the global decoder of $\mathrm{RS}_q[q,d]$ on the restriction of the file to ℓ.
        We have corrected all the erasures on that line, x included.
        $E = E \cup \{x\}$
        Modify V accordingly.
  until $E = \emptyset$
  return V

Fig. 5. Our global decoding algorithm for lifted Reed-Solomon codes.

We now study the fail cases of the global decoding algorithm. Let $\mathrm{Lift}_m(\mathrm{RS}_q[q,d])$ be a lifted RS code. For an erased position $s \in \mathbb{F}_q^m$ to be unrecoverable, it is necessary that each line going through s possesses at least d erasures. However, it is not sufficient. Indeed, suppose that there exists a line ℓ going through s with exactly d erasures. If there exists an erasure position s' on the line ℓ and a line ℓ' going through s' with at most $d-1$ erasures then the symbol erased at position s' can be recovered using the $\mathrm{RS}_q[q,d]$ decoder. Since s' lies on ℓ, this means that ℓ now contains only $d-1$ erasures and they all can be corrected, the one at s included.

In order to capture a set of unrecoverable erasures for our global decoding algorithm, we introduce the following property:

Definition 5. *Let $\mathbb{F}_q$ be a finite field and m, d be positive integers. We say that a set $S \subseteq \mathbb{F}_q^m$ is a d-cover set if S verifies the following property:*

$$\forall s \in S, \forall \text{ line } \ell \subseteq \mathbb{F}_q^m \text{ going through } s, |S \cap \ell| \geq d$$

Or equivalently, for all line $\ell \subseteq \mathbb{F}_q^m, |S \cap \ell| = 0$ or $|S \cap \ell| \geq d$

Since the d-cover subsets of $\mathbb{F}_q^m$ represent the unrecoverable erasure patterns, we want to find an audit procedure that can detect their existence with high probability and low communication complexity. We propose the following audit:

1. The client randomly chooses a line $\ell \subseteq \mathbb{F}_q^m$.
2. The client retrieves the restriction of the outsourced file to the chosen line.
3. If it contains d or more erasures, reject, if not, accept.

The next step is to determine the probability that this audit detects a set of unrecoverable erasures if one exists. Let $S \subseteq \mathbb{F}_q^m$ be a non-empty d-cover set. Then there exists $s \in S$ and for each line ℓ going through s, $|\ell \cap S| \geq d$. We also know that for any line $\ell \subseteq \mathbb{F}_q^m$, either $|\ell \cap S| = 0$ or $|\ell \cap S| \geq d$.

Recall that $L := (q^m - 1)/(q-1)$ is the number of lines going through a point in $\mathbb{F}_q^m$ and that $q^{m-1}L$ is the total number of lines in $\mathbb{F}_q^m$. Let ℓ be the randomly chosen line for the audit and s be an element of S. We have:

$$\Pr[|\ell \cap S| \neq 0] = \frac{L}{q^{m-1}L} + \left(1 - \frac{1}{q^{m-1}}\right) \cdot \Pr[|\ell \cap S| \neq 0 \mid s \notin \ell]$$

Let E be the event $\{|\ell \cap S| \neq 0 \mid s \notin \ell\}$. For each point $x \in \ell$, there is a unique line (xs) going through x and s. Since $s \in S$, this line contains at least d erased points in S, one being s. Since lines in $\mathbb{F}_q^m$ have q points, the probability that $x \in S$ is at least $(d-1)/(q-1)$. Moreover, if at least $q-d+1$ points of ℓ do not belong to S we immediately know that $\ell \cap S = \emptyset$ since, by definition of S, either $|\ell \cap S| = 0$ or $|\ell \cap S| \geq d$. Thus, $\Pr[E] \geq 1 - (1 - (d-1)/(q-1))^{q-d+1}$.

$$\text{Therefore,} \quad \Pr[|\ell \cap S| \neq 0] \geq 1 - \left(1 - \frac{1}{q^{m-1}}\right)\left(1 - \frac{d-1}{q-1}\right)^{q-d+1}.$$

The calculation we just made is essential. Indeed, since we supposed $S \neq \emptyset$, the event $\neg\{|\ell \cap S| \neq 0\}$ can be interpreted as 'on probed line ℓ, the audit accepts although the file is not retrievable'. In the CC security model for PoR, this is exactly the advantage of the distinguisher, *i.e.* the security of the scheme. In other words, we just upper-bounded the security of our PoR scheme.

We now formally prove the security of our PoR in the CC framework. We quickly describe the converters $\mathsf{lift\_rs\_por}_{\mathsf{init}}$ and $\mathsf{lift\_rs\_por}_{\mathsf{audit}}$. Both use the encoder and global decoder for lifted RS codes. On input $(\texttt{read}, i)$, both converters retrieve the whole memory using `read` requests, then they call the global decoder on the obtained word (corrupted values ϵ are replaced with erasures) and return the i-th symbol of the output if decoding succeeds. On input $(\texttt{write}, i, x)$, both converters retrieve the whole memory with read requests and decode it like before. If decoding succeeds, they replace the i-th symbol by x, encode the whole word and store it on the SMR using `write` requests.

On input `audit`, $\mathsf{lift\_rs\_por}_{\mathsf{audit}}$ chooses a random line $\ell \subseteq \mathbb{F}_q^m$ and retrieves the restriction of the outsourced file to ℓ through `read` requests. If the restriction contains d or more erasures, it returns `reject`. If not, it returns `accept`.

Theorem 1. *Let $d, m, \ell \in \mathbb{N}$, $\mathbb{F}_q$ be a finite field. The protocol* $\mathsf{lift\_rs\_por} := (\mathsf{lift\_rs\_por}_{\mathsf{init}}, \mathsf{lift\_rs\_por}_{\mathsf{audit}})$ *for the lifted RS code* $\mathrm{Lift}_m(\mathrm{RS}_q[q,d])$ *of dimension ℓ constructs the auditable and authentic SMR, say* $\mathbf{aSMR}^{\mathsf{audit}}_{\Sigma,\ell}$*, from* $\mathbf{aSMR}_{\Sigma,q^m}$*, with respect to the simulator* $\mathsf{sim}_{\mathsf{audit}}$*. More precisely, for all distinguishers* $\mathbf{D}$ *making at most r audits, we have*

$$\Delta^{\mathbf{D}}(\mathsf{lift\_rs\_por}_{\mathcal{P}}\, \mathbf{aSMR}_{\Sigma,q^m}, \mathsf{sim}^{\mathsf{S}}_{\mathsf{audit}}\, \mathbf{aSMR}^{\mathsf{audit}}_{\Sigma,\ell}) \leq r \cdot \left(1 - \frac{1}{q^{m-1}}\right)\left(1 - \frac{d-1}{q-1}\right)^{q-d+1}$$

Proof. Since our scheme is clearly correct (*i.e.* the client can always retrieve its file when there is no adversary), we alleviate notations and proofs by omitting correctness. We prove security by comparing the behaviors of the audit of the real system (the $\mathbf{aSMR}$ with the protocol) with that of the ideal one (the $\mathbf{aSMR}^{\mathsf{audit}}$ with the simulator). We describe the simulator $\mathsf{sim}_{\mathsf{auth}}$. It maintains a simulated memory, emulating the real world memory, using the history of the ideal resource. On $(\texttt{delete}, i)$, the simulator replaces the i-th entry of its simulated memory by ϵ. On $(\texttt{restore}, i)$, the simulator restores the content of the i-th entry of its simulated memory to the last value written there. The simulator maintains a simulated history using the ideal history of the $\mathbf{aSMR}^{\mathsf{audit}}$.

If, after a `delete` request, the set of corrupted locations of the simulated memory contains a d-cover subset of $\mathbb{F}_q^m$, the simulator deletes the whole ideal memory by sending `delete` requests to $\mathbf{aSMR}^{\mathsf{audit}}$. Similarly, if after a `restore` request, the set of corrupted locations of the simulated memory does not contain a d-cover subset of $\mathbb{F}_q^m$, the simulator restores the whole ideal memory by sending `restore` requests to $\mathbf{aSMR}^{\mathsf{audit}}$.

On an `audit` request, the simulator chooses a random line $\ell \subseteq \mathbb{F}_q^m$, adds the entries $(\texttt{read}, i)$ for $i \in \ell$ to its simulated history. Then, if the restriction of its simulated memory to ℓ contains strictly less than d corrupted cells, the simulator sends `allow` to $\mathbf{aSMR}^{\mathsf{audit}}$. Else, it instructs the $\mathbf{aSMR}^{\mathsf{audit}}$ to output `reject`.

Upon `auditReq` at interface S_H : Recall that d-cover sets are the sets of unrecoverable erasures for our global decoder of lifted RS codes. Suppose that a subset of the corrupted cells forms a d-cover set. In order to run the audit, the converter chooses a random line $\ell \subseteq \mathbb{F}_q^m$, retrieves the restriction of the memory to this line through `read` requests and adds the corresponding entries to its simulated history. We showed, see equation 5.2, that the probability that this restriction contains strictly less than d erasures, *i.e.*, that the audit is successful, is less than $(1 - 1/q^{m-1})(1 - (d-1)/(q-1))^{q-d+1}$.

The simulation is perfect unless the following BAD event occurs: *having simulated a real audit, the simulator answers* `allow` *(audit should succeed) whereas a d-cover subset of corrupted cells exists.* In that case, the simulator has chosen a restriction of the memory to a line ℓ that contains strictly less than d corrupted cells, and has written the corresponding `read` requests to its simulated history. Note that the distinguisher has access to the simulated history. Then, the simulator outputs `allow` to the ideal resource, that runs the ideal audit. Since there exists a d-cover set of corrupted memory cells, the file is unretrievable so the ideal audit fails and the client receives `reject`. The distinguisher thus observes the following incoherence: `reject` is output while the (simulated) history contains the trace of a valid audit. The adversary knows that this is the ideal system.

To sum up, the only observable difference from a distinguisher point of view lies in the audit procedure. The overall distinguishing probability is thus the one of distinguishing a real audit from a simulated one. As we saw, if the distinguisher runs r audits, this probability is less than $r \cdot (1-1/q^{m-1}) \cdot (1-(d-1)/(q-1))^{q-d+1}$, yielding the aforementioned result. □

5.3 The graph code PoR scheme

We give another instantiation of our framework using the graph codes of Tanner [16]. We briefly recall how these codes are constructed.

Let $G = (V, E)$ be a q-regular graph on n vertices. For a vertex $v \in V$, let $\Gamma(v)$ be the set of vertices adjacent to v. Let $\mathbb{F}$ be a finite field and let $\mathcal{C}_0 \subseteq \mathbb{F}^q$ be a linear code, called the *inner* code. Fix an arbitrary order on the edges incident to each vertex of G and let $\Gamma_i(v)$ be the i-th neighbor of v. A Tanner code is defined as the set of all labelings of the edges of G that respect the inner code $\mathcal{C}_0$. Formally,

Definition 6 (Tanner code). *Let $G = (V, E)$ be a q-regular graph on n vertices and let $\mathcal{C}_0 \subseteq \mathbb{F}^q$ be a linear code. The Tanner code $\mathcal{C}(G, \mathcal{C}_0) \subseteq \mathbb{F}^E$ is a linear code of length $nq/2$, so that for $c \in \mathbb{F}^E$, $c \in \mathcal{C}(G, \mathcal{C}_0)$ if and only if, for all $v \in V$, $(c_{(v,\Gamma_1(v))}, \ldots, c_{(v,\Gamma_q(v))}) \in \mathcal{C}_0$.*

One can easily check, by counting constraints, that if $\mathcal{C}_0$ has rate R, then $\mathcal{C}(G, \mathcal{C}_0)$ has rate at least $2R-1$. These codes possess some sort of local correction. Indeed, to correct an edge e incident to a vertex v, one can retrieve the vector $(c_{(v,\Gamma_1(v))}, \ldots, c_{(v,\Gamma_q(v))})$ of labels of edges incident to v and correct it using the decoder of $\mathcal{C}_0$.

In the following, let d be the minimum distance of the inner code. Again, using the composability of CC, we only have to deal with potential erasures. Following our framework of sec. 3, we start by sketching our global decoder. In the following, we say that an edge is *erased* when the label of that edge is erased. Similarly, we say that we *correct* an edge if we correct the label of that edge.

Assume that we want to correct an erasure on an edge e incident to a vertex v. If v is incident to less than $d-1$ erased edges, we can use the erasure decoding of $\mathcal{C}_0$ to correct all the edges incident to v, e included. Otherwise, v is incident to $k > d-1$ erased edges. Pick an erased edge incident to v. This edge is also incident to a vertex $v' \neq v$. If v' is incident to less than $d-1$ erased edges, we can correct them all and v is now incident to $k-1$ erased edges. If $k-1 \leq d-1$ we can correct the edge e. Else, we iterate the process on v and its neighbors.

Now, we have to characterize the configurations of erased edges that are unrecoverable for our decoding algorithm. We claim that these unrecoverable configurations correspond to subgraphs of G of minimum degree d. Indeed, these are the graph analogues of the d-cover sets for lifted RS codes. We prove our claim: suppose that the subgraph formed by the unrecoverable edges possesses a vertex v incident to less than $d-1$ unrecoverable edges. Then, by iterating the local decoding algorithm, we can recover the other edges incident to v so that only these unrecoverable edges remain erased. Then, since there are less than $d-1$ erased edges incident to v and since the minimum distance of the inner code is d, we can correct all the edges incident to v using the decoder of the inner code. This is in contradiction with these edges being unrecoverable.

Finally, the audit consists in randomly choosing a vertex v and retrieving the vector $w := (c_{(v,\Gamma_1(v))}, \ldots, c_{(v,\Gamma_q(v))})$ of labeling of edges incident to v. If w contains d or more erasures, the audit rejects. Else, it accepts.

The security of the audit depends on the graph G and the minimum distance d of the inner code $\mathcal{C}_0$. The bigger the minimum subgraphs of G with minimum degree d are, the better the security of the PoR will be. Indeed, let s be the minimum size (number of vertices) of a subgraph of G with minimum degree d. For a configuration of unrecoverable erasures to exist, we thus need at least s vertices of G with at least d erased edges. Recall that our audit chooses a random vertex of G and accepts if and only if this vertex is incident to less than $d-1$ erased edges. Thus, the probability that our audit accepts when there exists an unrecoverable set of erased edges is less than $1 - s/|V|$. In our framework, this is exactly the advantage of the adversary in breaking the security of our PoR. A similar proof ans simulator to the ones of th. 1 yield the following theorem:

Theorem 2. *Let $G = (V, E)$ be a q-regular graph with $|V| = n$ and let $\mathcal{C}_0 \subseteq \mathbb{F}^q$ be a linear code with minimum distance d and rate R. Let s be the minimum size (number of vertices) of a subgraph of G with minimum degree d. The protocol* $\mathsf{graph\_por} := (\mathsf{graph\_por}_{\mathsf{init}}, \mathsf{graph\_por}_{\mathsf{audit}})$ *for a Tanner code $\mathcal{C}(G, \mathcal{C}_0)$ of length $nq/2$ and rate at least $2R-1$ that:*

1. *Starts by encoding the file and uploads it to the server.*
2. *On an* `audit` *request, chooses a random vertex $v \in V$ and accepts if and only if v is incident to less than $d-1$ erased edges.*

3. Extracts the file using the algorithm sketched above.

constructs the auditable and authentic SMR, say $\mathbf{aSMR}^{\mathsf{audit}}_{\mathbb{F},(2R-1)nq/2}$*, from* $\mathbf{aSMR}_{\mathbb{F},nq/2}$*, with respect to the simulator* $\mathsf{sim}_{\mathsf{audit}}$*. More precisely, for all distinguishers* $\mathbf{D}$ *making at most* r *audits, we have*

$$\Delta^{\mathbf{D}}(\mathsf{graph\_por}_{\mathcal{P}}\,\mathbf{aSMR}_{\mathbb{F},nq/2}, \mathsf{sim}^{\mathsf{S}}_{\mathsf{audit}}\,\mathbf{aSMR}^{\mathsf{audit}}_{\mathbb{F},(2R-1)nq/2}) \leq r \cdot \left(1 - \frac{s}{n}\right)$$

6 Parameters

The impact of the choice of the lifted RS code on the parameters of our lifted RS PoR scheme are highlighted in fig. 7. The grey line gives a choice of parameters with a storage overhead of 13.9% and total communication of 0.01% of the file size. Increasing the length q of the RS base code decreases the storage overhead and increasing the lifting parameter m increases the size of the file stored. Exact formulae for the parameters of our scheme is given in fig. 6.

Let us compare our parameters with the ones of [11]. First, in both schemes, the client's file is encoded using a lifted RS code and the audit consists in probing the restriction of this codeword to a random affine line. In our case, we authenticate the data using our MAC based authentication protocol (see sec. 4) whereas [11] binds data to a specific location by using an encryption scheme. Let κ be the computational security parameter of both schemes and Σ be the alphabet of the code. Our scheme stores a code symbol along with a MAC tag, that is $\kappa + \log|\Sigma|$ bits, in each memory location of the server whereas [11] stores a ciphertext, that is κ bits, in each memory location. Since $\log|\Sigma| \ll \kappa$ (we have $\kappa = 128$ and $|\Sigma| = q$ in fig. 7), our scheme and the one of [11] have very close storage overhead and communication complexity. In [11], the minimum distance of the code d is chosen to be equal to 2. Using our security analysis of th. 1, we show that the [11] scheme has only 1.44 bits of statistical security, when $d = 2$, whereas state-of-the-art schemes expect at least 40. See fig. 7 for our recommended parameters.

A major benefit of our scheme is that our audit produces less "false positives" than the one of [11]. For PoRs, a false positive occurs when an audit rejects while the file is still retrievable. In other words, the client thinks that he lost his file, but it is still retrievable in full. The number of false positive audits has no influence on the security of the PoR but, in practice, it is a situation that we absolutely wish to avoid. The audit of [11] rejects if the restriction of the file to an affine line does not belong to the RS base code. In other words, if there is at least one corruption on the line probed by the audit, it deems the file unretrievable. If the adversary introduces at least one erasure on every line of the space, the audit would always reject independently of the correction capability (*i.e.* the minimum distance) of the code. Using our framework and our authentication protocol, we are able to fix this problem. Indeed, our audit deems the file unretrievable only if the probed line contains at least d erasures, where d is the minimum distance of the RS base code. This means that we drastically decrease the number of false positive audits, making our scheme much more reliable and usable in practice.

For example, suppose that the outsourced file is encoded using a lifted RS code over $\mathbb{F}_q^2$ with minimum distance $d \geq 3$. Let ℓ_1, ℓ_2 be two intersecting lines

in $\mathbb{F}_q^2$. Suppose that an adversary erases all the file's symbols at the locations given by ℓ_1 and ℓ_2 and no other symbols. Of course, the file is still retrievable since the local decoder can correct all the erasures of $\ell_1 \setminus \ell_2$ by querying all the lines parallel to ℓ_2 (these lines contain only one erasure and $d \geq 3$). Then, the local decoder can correct all the erasures of ℓ_2 by querying any line intersecting ℓ_2. Unfortunately, in this situation, the audit of [11] rejects with probability 1. Indeed, their audit chooses a random line ℓ in $\mathbb{F}_q^2$ and rejects if ℓ contains at least one erasure. This is always the case here since, either ℓ intersects ℓ_1 or, ℓ is parallel to ℓ_1 and is thus intersecting ℓ_2. This is not the case with our audit. Indeed, since only two lines of $\mathbb{F}_q^2$ have d or more erasures, our audit rejects with probability $2/(q^2+q)$ since there are q^2+q lines in $\mathbb{F}_q^2$.

Future work includes evaluating the efficiency of our Tanner code PoRs according to different choices of inner codes and graphs as well as instantiating our framework with other families of high-rate locally correctable codes.

	Exact value	Asymptotics
C. storage overhead	κ	$\mathcal{O}(1)$
S. storage overhead	$(\frac{1}{R}-1)\lvert F\rvert + q^m\kappa$	$\mathcal{O}(\lvert F\rvert)$
C. $\rightarrow$ S.	$2m\log q$	$\mathcal{O}(\lvert F\rvert)$
S. $\rightarrow$ C.	$q(\kappa+\log q)$	$\mathcal{O}(\lvert F\rvert^{1/m})$

Fig. 6. The exact parameters of our scheme. $|F|$ denotes the file size in bits, κ the security parameter of the MAC, q the field size and $m \geq 2$ the lifting parameter. We have $Rq^m \log q = |F|$.

PoR param.			Results					
m	q	d	$\lvert F\rvert$ (bits)	$\frac{1}{R}-1$	comm. C. $\rightarrow$ S. (bits)	comm. S. $\rightarrow$ C. (bits)	comm./$\lvert F\rvert$	Statistical Security
2	256	32	255003	1.056	32	2048	0.0081	2^{-42}
	512		1446533	0.631	36	4608	0.0032	2^{-43}
	1024		7441987	0.409	40	10240	0.0013	2^{-44}
	2048		36072982	0.279	44	22528	0.0006	2^{-44}
	4096		168474135	0.195	48	49152	0.0003	2^{-44}
	8192		765948403	0.139	52	106496	0.0001	2^{-44}
	1024	64	6389859	0.641	40	10240	0.0016	2^{-88}
	2048		32605896	0.415	44	22528	0.0007	2^{-89}
	4096		157041023	0.282	48	49152	0.0003	2^{-90}
	8192		728834780	0.197	52	106496	0.0001	2^{-90}

Fig. 7. Different choices of lifted Reed-Solomon codes for our PoR scheme.

References

1. Badertscher, C., Maurer, U.: Composable and robust outsourced storage. In: Topics in Cryptology - CT-RSA 2018 - The Cryptographers' Track at the RSA Conference 2018, San Francisco, CA, USA, April 16-20, 2018, Proceedings. pp. 354–373 (2018). https://doi.org/10.1007/978-3-319-76953-0_19
2. Bowers, K.D., Juels, A., Oprea, A.: Proofs of retrievability: Theory and implementation. In: Proceedings of the 2009 ACM Workshop on Cloud Com-

puting Security. pp. 43–54. CCSW '09, ACM, New York, NY, USA (2009). https://doi.org/10.1145/1655008.1655015
3. Canetti, R.: Universally composable security: a new paradigm for cryptographic protocols. In: Proceedings 42nd IEEE Symposium on Foundations of Computer Science. p. nil (- 2001), https://doi.org/10.1109/sfcs.2001.959888
4. Dodis, Y., Vadhan, S., Wichs, D.: Proofs of retrievability via hardness amplification. In: Proceedings of the 6th Theory of Cryptography Conference on Theory of Cryptography. pp. 109–127. TCC '09, Springer-Verlag, Berlin, Heidelberg (2009). https://doi.org/10.1007/978-3-642-00457-5_8
5. Guo, A., Kopparty, S., Sudan, M.: New affine-invariant codes from lifting. In: Proceedings of the 4th Conference on Innovations in Theoretical Computer Science. pp. 529–540. ITCS '13, ACM, New York, NY, USA (2013). https://doi.org/10.1145/2422436.2422494
6. Hemenway, B., Ostrovsky, R., Wootters, M.: Local correctability of expander codes. In: Fomin, F.V., Freivalds, R., Kwiatkowska, M., Peleg, D. (eds.) Automata, Languages, and Programming. pp. 540–551. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
7. Jost, D., Maurer, U.: Overcoming Impossibility Results in Composable Security Using Interval-Wise Guarantees, pp. 33–62. Advances in Cryptology - CRYPTO 2020, Springer International Publishing (2020)
8. Juels, A., Kaliski, Jr., B.S.: Pors: Proofs of retrievability for large files. In: Proceedings of the 14th ACM Conference on Computer and Communications Security. pp. 584–597. CCS '07, ACM, New York, NY, USA (2007). https://doi.org/10.1145/1315245.1315317
9. Katz, J., Trevisan, L.: On the efficiency of local decoding procedures for error-correcting codes. In: Proceedings of the Thirty-second Annual ACM Symposium on Theory of Computing. pp. 80–86. STOC '00, ACM, New York, NY, USA (2000). https://doi.org/10.1145/335305.335315
10. Kopparty, S., Saraf, S., Yekhanin, S.: High-rate codes with sublinear-time decoding. In: Proceedings of the Forty-third Annual ACM Symposium on Theory of Computing. pp. 167–176. STOC '11, ACM, New York, NY, USA (2011). https://doi.org/10.1145/1993636.1993660
11. Lavauzelle, J., Levy-Dit-Vehel, F.: New proofs of retrievability using locally decodable codes. In: International Symposium on Information Theory ISIT 2016. pp. 1809 – 1813. Barcelona, Spain (2016). https://doi.org/10.1109/ISIT.2016.7541611
12. Maurer, U.: Constructive Cryptography - A New Paradigm for Security Definitions and Proofs, pp. 33–56. Theory of Security and Applications, Springer Berlin Heidelberg (2012)
13. Maurer, U., Renner, R.: Abstract cryptography. In: In Innovations In Computer Science. Tsinghua University Press (2011)
14. Paterson, M., Stinson, D., Upadhyay, J.: A coding theory foundation for the analysis of general unconditionally secure proof-of-retrievability schemes for cloud storage. Journal of Mathematical Cryptology **7**(3), 183–216 (2013). https://doi.org/doi:10.1515/jmc-2013-5002
15. Shacham, H., Waters, B.: Compact proofs of retrievability. In: Pieprzyk, J. (ed.) Advances in Cryptology - ASIACRYPT 2008. pp. 90–107. Springer Berlin Heidelberg, Berlin, Heidelberg (2008)
16. Tanner, R.: A recursive approach to low complexity codes. IEEE Transactions on Information Theory **27**(5), 533–547 (1981). https://doi.org/10.1109/TIT.1981.1056404