Combinatorial techniques for repairing shares in threshold schemes

Douglas R. Stinson

David R. Cheriton School of Computer Science
University of Waterloo

Combinatorics 2018

Arco, Italy, June 4, 2018

Secret Sharing

- Various types of shared control schemes depend on a cryptographic primitive called a **(t, n)-threshold scheme**.
- Let t and n be positive integers, where $t \leq n$.
- The value t is the **threshold**.
- There is a trusted authority, denoted **dealer**, and n **users**, denoted $\boldsymbol{U_1}, \ldots, \boldsymbol{U_n}$.
- The dealer has a secret value $\boldsymbol{K} \in \mathcal{K}$, called a **secret** or a **key**, where $\mathcal{K}$ is a specified finite set.

Secret Sharing

- The dealer uses a **share generation algorithm** to split K into n **shares**, denoted $\boldsymbol{s_1}, \ldots, \boldsymbol{s_n}$.
- Each share $s_i \in \mathcal{S}$, where $\mathcal{S}$ is a specified finite **share set**.
- For every i, $1 \leq i \leq n$, the share s_i is transmitted by the dealer to user U_i using a secure channel.
- The following two properties should hold:
    1. a **reconstruction algorithm** can be used to reconstruct the secret, given any t of the n shares,
    2. no $t-1$ shares reveal any information as to the value of the secret.

An (n, n)-Threshold Scheme

- Suppose $K \in \mathbb{Z}_m$ is the secret.
- Let $s_1, \ldots, s_{n-1}$ be chosen **independently and uniformly at random** from $\mathbb{Z}_m$.
- Let

$$s_n = K - \sum_{i=1}^{n-1} s_i \bmod m.$$

- $s_1, \ldots, s_n$ are shares of an **(n, n)-threshold scheme**:
  1. the secret is reconstructed using the formula

     $$K = \sum_{i=1}^{n} s_i \bmod m,$$

     and
  2. given all the shares except s_j, K could take on any value, depending on the value of the "missing" share, s_j.

Shamir Threshold Scheme

- In 1979, Shamir showed how to construct a **(t, n)-threshold scheme** based on **polynomial interpolation** over $\mathbb{Z}_p$, where p is prime.
- This is really a **Reed-Solomon code** in disguise.
- Let $p \geq n + 1$ be a prime.
- Let $\mathcal{K} = \mathcal{S} = \mathbb{Z}_p$.
- In an **initialization phase**, $x_1, x_2, \ldots, x_n$ are defined to be n distinct non-zero elements of $\mathbb{Z}_p$.
- the dealer gives x_i to U_i, for all i, $1 \leq i \leq n$.
- The x_i's are **public** information.

Share Generation

Protocol: *Shamir threshold scheme share generation*

Input: A secret $K \in \mathbb{Z}_p$.

1. The dealer chooses $a_1, \ldots, a_{t-1}$ **independently and uniformly at random** from $\mathbb{Z}_p$.
2. The dealer defines

$$a(x) = K + \sum_{j=1}^{t-1} a_j \, x^j$$

   (note that $a(x) \in \mathbb{Z}_p[x]$ is a **random polynomial** of degree at most $t-1$, such that the **constant term** is the secret, K).
3. For $1 \leq i \leq n$, the dealer constructs the share $s_i = a(x_i)$ and gives it to U_i using a secure channel.

Reconstruction

- Suppose t users, say $U_{i_1}, \ldots, U_{i_t}$, want to determine K.
- They know that $s_{i_j} = a(x_{i_j})$, $1 \leq j \leq t$.
- Since $a(x)$ is a polynomial of degree at most $t-1$, they can determine $a(x)$ by **Lagrange interpolation**; then $K = a(0)$.
- The **Lagrange interpolation formula** is as follows:

$$a(x) = \sum_{j=1}^{t} s_{i_j} \prod_{1 \leq k \leq t, k \neq j} \frac{x - x_{i_k}}{x_{i_j} - x_{i_k}}.$$

- set $x = 0$; then

$$\begin{aligned} K &= \sum_{j=1}^{t} s_{i_j} \prod_{1 \leq k \leq t, k \neq j} \frac{-x_{i_k}}{x_{i_j} - x_{i_k}} \\ &= \sum_{j=1}^{t} s_{i_j} \prod_{1 \leq k \leq t, k \neq j} \frac{x_{i_k}}{x_{i_k} - x_{i_j}}. \end{aligned}$$

Reconstruction (cont.)

Protocol: *Shamir scheme secret reconstruction*

Input: $x_{i_1}, \ldots, x_{i_t}, s_{i_1}, \ldots, s_{i_t}$

1. For $1 \leq j \leq t$, define the **Lagrange coefficients**

$$\boldsymbol{b_j = \prod_{1 \leq k \leq t, k \neq j} \frac{x_{i_k}}{x_{i_k} - x_{i_j}}}.$$

   Note: the b_j's do not depend on the shares, so they can be **precomputed** (for a given subset of t users).

2. Compute

$$\boldsymbol{K = \sum_{j=1}^{t} b_j \, s_{i_j}}.$$

Example

- Suppose that $p = 17$, $t = 3$, and $n = 5$; and the public x-co-ordinates are $x_i = i$, $1 \leq i \leq 5$.
- Suppose that the users U_1, U_3, U_5 wish to compute K, given their shares $\mathbf{8}$, $\mathbf{10}$ and $\mathbf{11}$, respectively.
- The following computations are performed:

$$
\begin{aligned}
b_1 &= \frac{x_3 x_5}{(x_3 - x_1)(x_5 - x_1)} \bmod 17 \\
&= 3 \times 5 \times (2)^{-1} \times (4)^{-1} \bmod 17 \\
&= \mathbf{4}, \\
b_2 &= \mathbf{3}, \quad \text{and} \\
b_3 &= \mathbf{11} \\
K &= \mathbf{4} \times \mathbf{8} + \mathbf{3} \times \mathbf{10} + \mathbf{11} \times \mathbf{11} \bmod 17 = 13.
\end{aligned}
$$

Security of the Shamir Scheme

- Suppose $t-1$ users, say $U_{i_1}, \ldots, U_{i_{t-1}}$, want to determine K.
- They know that $s_{i_j} = a(x_{i_j})$, $1 \leq j \leq t-1$.
- Let K_0 be arbitrary.
- By **Lagrange interpolation**, there is a unique polynomial $a_0(x)$ such that

$$s_{i_j} = a_0(x_{i_j})$$

for $1 \leq j \leq t-1$ and such that

$$K_0 = a_0(0).$$

- Hence **no value of K can be ruled out**, given the shares held by $t-1$ users.

Security of the Shamir Scheme (cont.)

- With a bit more work, we can show that the **Shamir scheme** satisfies a property analogous to **perfect secrecy**.
- We assume an arbitrary but fixed **a priori** probability distribution on $\mathcal{K}$.
- Given any set of $\tau \leq t-1$ or fewer shares, say s_{i_j}, $j = 1, \ldots, \tau$, and given any $K_0 \in \mathcal{K}$, it is possible to show that

$$\text{Prob}[K = K_0 | s_{i_1}, \ldots, s_{i_\tau}] = \text{Prob}[K = K_0].$$

Repairability

- Suppose that a user U_ℓ (in a **(t, n)-threshold scheme**, say) loses their share.
- The goal is to find a **secure protocol**, involving U_ℓ and a subset of the other users, that allows the missing share s_ℓ to be reconstructed.
- We are considering a setting where the dealer is **no longer present** in the scheme after the initial setup.
- We will assume secure pairwise channels linking pairs of users.
- Three techniques for repairing shares:
  1. the **enrollment scheme** (Nojoumian [3])
  2. **secure regenerating codes** (Shah, Rashmi and Kumar [4])
  3. **combinatorial schemes** (Stinson and Wei [5])
- For a survey of these techniques, see Laing and Stinson [2].

Repairability (cont.)

A **(t, n, d)-repairable threshold scheme**, which we abbreviate to **(t, n, d)-RTS**, is a protocol that operates in two phases:

1. In the **message exchange phase**, a certain subset of d users (not including P_ℓ) exchange messages among themselves. The integer d is called the **repairing degree**. We will only consider protocols where each user sends at most one message to any other user, and every message is sent at the same time.
2. In the **repairing phase**, these same d users each send a message to P_ℓ. The messages received by P_ℓ allow P_ℓ's share to be reconstructed. Some of the protocols we study only require a repairing phase.

We note that $d \geq t$ is an obvious **necessary condition** for the existence of such a scheme. (WHY?)

Enrollment Protocol

- The *Enrollment Protocol* is a (t, n, t)**-RTS** that is based on a (t, n)**-Shamir threshold scheme**.
- Suppose that users $U_1, \ldots, U_t$ want to repair the share for user U_ℓ, where $\ell > t$.
- The share for P_ℓ is $s_\ell = a(\ell)$.
- From the **Lagrange Interpolation Formula**, setting $x = x_\ell$, the share s_ℓ can be expressed as

$$s_\ell = \sum_{i=1}^{t} b_i s_i,$$

where the b_i's are public Lagrange coefficients.

Enrollment Protocol (cont.)

Message-exchange phase

1. For all $1 \leq i \leq t$, user U_i splits the "secret" $b_i s_i$ into t shares using a (t, t)-threshold scheme:

$$\boldsymbol{b_i s_i = \sum_{j=1}^{t} \delta_{j,i}}.$$

2. Then, for all i, j, user U_i transmits $\delta_{j,i}$ to user U_j.

Repairing phase

1. For all j, user U_j transmits σ_j to user U_ℓ, where

$$\boldsymbol{\sigma_j = \sum_{i=1}^{t} \delta_{j,i}}.$$

2. Finally, user U_ℓ computes their share s_ℓ using the formula

$$\boldsymbol{s_\ell = \sum_{j=1}^{t} \sigma_j}.$$

Share-exchange Matrix

It is convenient to consider the following **share-exchange matrix**:

$$\mathcal{E} = \begin{pmatrix} \delta_{1,1} & \delta_{2,1} & \cdots & \delta_{t,1} \\ \delta_{1,2} & \delta_{2,2} & \cdots & \delta_{t,2} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{1,t} & \delta_{2,t} & \cdots & \delta_{t,t}. \end{pmatrix}.$$

- The sum of **the entries in the ith row** of $\mathcal{E}$ is equal to $b_i s_i$.
- The sum of **the entries in the jth column** of $\mathcal{E}$ is equal to σ_j.
- The sum of **all the entries** in $\mathcal{E}$ is equal to s_ℓ.
- U_ℓ is given the t column sums, so U_ℓ can compute s_ℓ.

Comments and Properties of the Enrollment Protocol

- The basic technique goes back to early studies on **secure multiparty computation** from the 1980s.
- We have **universal repairability**: any set of t users can repair any other share.
- The protocol is secure against **honest-but-curious** coalitions of size $t-1$.
- The number of messages sent during the protocol, namely $\boldsymbol{t^2}$, is quadratic in t, which could be considered a drawback of the scheme.
- An improved version is described in [2], in which user U_i does not send a message to user U_j if $j > i$. This modification is still secure, and it achieves **optimal** communication complexity $\boldsymbol{t(t+1)/2}$.

A $(2,5,3)$-RTS based on a Regenerating Code (Example)

- There are five components to a **message**: $K_1, \ldots, K_5$.
- Three components are **random** and the other two components comprise the **secret**.
- There are $\boldsymbol{n = 5}$ users.
- *Share Generation:* Each user is given a share consisting of three components, where each component is a certain linear combination of the K_i's.
- Any user U_j can repair their share with information provided by any $\boldsymbol{d = 3}$ other "helper" users.
- The shares belonging to any $\boldsymbol{t = 2}$ users yield a a system of linear equations that can be solved to obtain the entire message $K_1, \ldots, K_5$.
- Thus they can obtain the secret.
- It can also be proven that no $\boldsymbol{t - 1 = 1}$ user can compute any information about the secret.

Combinatorial RTS

- As an example, we construct a $(2, 12, 3)$**-RTS**.
- Start with a $(9, 3, 1)$**-BIBD** (an affine plane of order 3), which has **12 blocks**. This is the **distribution design**.
- We associate a block of the design with each user:

$$\begin{array}{ccc} U_1 \leftarrow \{1,2,3\} & U_2 \leftarrow \{4,5,6\} & U_3 \leftarrow \{7,8,9\} \\ U_4 \leftarrow \{1,4,7\} & U_5 \leftarrow \{2,5,8\} & U_6 \leftarrow \{3,6,9\} \\ U_7 \leftarrow \{1,5,9\} & U_8 \leftarrow \{2,6,7\} & U_9 \leftarrow \{3,4,8\} \\ U_{10} \leftarrow \{1,6,8\} & U_{11} \leftarrow \{2,4,9\} & U_{12} \leftarrow \{3,5,7\} \end{array}$$

- Each user gets three shares from a $(5, 9)$**-threshold scheme** (the **base scheme**), as specified by the associated block.
- Each share in the resulting RTS consists of three **subshares**.
- Any **two blocks** of the distribution design contain **at least five points**, whereas **one block** contains only **three points**.
- Therefore **two users** can reconstruct the secret, but **one user** cannot (since the base scheme has threshold 5).

Repairability (Example)

- When a user wants to repair their share, they contact **three other users** who have the relevant subshares.
- For example, U_1 could contact U_4 to obtain subshare #1, U_8 to obtain subshare # 2 and U_{12} to obtain subshare #3:

$$\begin{array}{ccc} U_1 \leftarrow \{\mathbf{1}, \mathbf{2}, \mathbf{3}\} & U_2 \leftarrow \{4, 5, 6\} & U_3 \leftarrow \{7, 8, 9\} \\ U_4 \leftarrow \{\mathbf{1}, 4, 7\} & U_5 \leftarrow \{2, 5, 8\} & U_6 \leftarrow \{3, 6, 9\} \\ U_7 \leftarrow \{1, 5, 9\} & U_8 \leftarrow \{\mathbf{2}, 6, 7\} & U_9 \leftarrow \{3, 4, 8\} \\ U_{10} \leftarrow \{1, 6, 8\} & U_{11} \leftarrow \{2, 4, 9\} & U_{12} \leftarrow \{\mathbf{3}, 5, 7\} \end{array}$$

- We do not need to use all twelve blocks in the distribution design; for repairability, it suffices to have a subset of blocks such that each point is a **contained in at least two blocks**.
- We can take the **first six blocks**, along with **any subset of the last six blocks**, to construct a $\mathbf{(2, m, 3)}$**-RTS** for any $m \in \{6, \ldots, 12\}$.

Required Properties of a Distribution Design

1. In order to be able to construct a threshold scheme with threshold t, the distribution design must satisfy the property that **the number of points in the union of any t blocks is greater than the number of points in the union of any $t-1$ blocks**.

   Remark: This property implies that the distribution design is a t-cover free family.

2. In order to provide repairability for a variable number of users, we need to identify a small **basic repairing set**, which is a set of blocks in the design such that every point is contained in at least two of these blocks.

   Remark: Taking two **parallel classes** from a **resolvable design** will yield a basic repairing set of minimum possible size.

Projective Planes as Distribution Designs

Lemma 1

The union of any $t-1$ blocks (lines) in a projective plane of order q contain **at most** $\boldsymbol{q(t-1)+1}$ *points.*

Proof.

Denote the $t-1$ lines by $A_0, \ldots, A_{t-2}$. Each A_i $(i \geq 1)$ contains a point in A_0, so

$$\left|\bigcup_{i=0}^{t-2} A_i\right| \leq q+1+(t-2)q = q(t-1)+1.$$

□

Remark: Equality occurs if and only if the $t-1$ lines all contain a common point.

Projective Planes as Distribution Designs (cont.)

Lemma 2

For $t \leq q+1$, the union of any t lines in a projective plane of order q contain **at least** $\boldsymbol{t(q+1-(t-1)/2)}$ *points.*

Proof.

Denote the t lines by $A_0, \ldots, A_{t-1}$. Each A_i contains $q+1-i$ points that are not in $\bigcup_{h=0}^{i-1} A_h$. It follows that

$$\left|\bigcup_{i=0}^{t-1} A_i\right| \geq \sum_{i=0}^{t-1}(q+1-i) = t(q+1) - \frac{t(t-1)}{2}.$$

□

Remark: Equality occurs if and only if no three of the t lines are collinear, so they form the dual of a **$\boldsymbol{t}$-arc**.

Example

- Consider a projective plane of order 5.
- One block contains 6 points.
- Two blocks contain 11 points.
- Three blocks contain **at least** 15 and **at most** 16 points.
- Four blocks contain **at least** 18 and **at most** 21 points.
- Five blocks contain **at least** 20 points.
- We can accommodate thresholds 2 (since $\mathbf{6 < 11}$), 3 (since $\mathbf{11 < 15}$) and 4 (since $16 < 18$), but not 5 (since $\mathbf{21 \geq 20}$).

Basic Repairing Sets in Projective Planes

- Recall that a basic repairing set is a subset of blocks (lines) that contains **every point at least twice**.
- In the context of a projective plane, this is precisely the dual of a **2-blocking set** (see, e.g., Ball and Blokhuis [1]).
- A simple construction: Choose any three noncollinear points x, y and z of the projective plane, and take all the lines that contain at least one of these points. This yields a basic repairing set of size $\mathbf{3q}$.
- Another construction: Suppose that q is a square of a prime power. Start with two disjoint Baer subplanes in $\mathbf{PG(2, q)}$ and take all the lines that contain a line from either of these two subplanes. This yields a basic repairing set of size $\mathbf{2(q + \sqrt{q} + 1)}$, which is an improvement asymptotically over the previous construction.

Ramp Schemes

- A basic property of a (t, n)**-threshold scheme** is that $|\mathcal{K}| \leq |\mathcal{S}|$.
- In the Shamir threshold scheme, we have $|\mathcal{K}| = |\mathcal{S}|$.
- A **weaker** security property allows for **larger** secrets to be accommodated using the same size shares.
- In a (t_1, t_2, n)**-ramp scheme**, any t_2 shares permit reconstruction of the secret, but no information about the secret is revealed by any t_1 shares.
- If $t_1 = t_2 - 1$ we have a threshold scheme.
- In a (t_1, t_2, n)**-ramp scheme**, it holds that $|\mathcal{K}| \leq |\mathcal{S}|^{t_2 - t_1}$.

Construction of Ramp Schemes

A straightforward modification of the Shamir threshold scheme permits the construction of ramp schemes where this bound is met with equality.

Protocol: *Shamir ramp scheme share generation*

Input: A secret $K \in (\mathbb{Z}_p)^{t_2-t_1}$, say $K = (a_0, \ldots, a_{t_2-t_1-1})$.

1. The dealer chooses $a_{t_2-t_1}, \ldots, a_{t_2-1}$ **independently and uniformly at random** from $\mathbb{Z}_p$.
2. The dealer defines

$$\boldsymbol{a(x) = \sum_{j=0}^{t_2-1} a_j\, x^j}$$

3. For $1 \leq i \leq n$, the dealer constructs the share $\boldsymbol{s_i = a(x_i)}$ and gives it to U_i using a secure channel.

Ramp Schemes and Distribution Designs

- Suppose $\ell_1 < \ell_2$ and our distribution design satisfies the following two properties:
  - **the union of any $t-1$ blocks contains at most ℓ_1 points**
  - **the union of any t blocks contains at least ℓ_2 points**
- Then we can share a secret using a base scheme which is an **(ℓ_1, ℓ_2, m)-ramp scheme**, where m is the number of points in the distribution design.
- Previously, we were using an **(ℓ_2, m)-threshold scheme**.
- Using a ramp scheme allows the secret to be $\ell_2 - \ell_1$ times larger than before.

Example

- Consider a projective plane of order 5. As we already noted:
  - One block contains 6 points.
  - Two blocks contain 11 points.
  - Three blocks contain **at least** 15 and **at most** 16 points.
  - Four blocks contain **at least** 18 points.
- Therefore
  - for $t = 2$, we can take $\ell_1 = 6$, $\ell_2 = 11$, so $\boldsymbol{\ell_2 - \ell_1 = 5}$.
  - for $t = 3$, we can take $\ell_1 = 11$, $\ell_2 = 15$, so $\boldsymbol{\ell_2 - \ell_1 = 4}$.
  - for $t = 4$, we can take $\ell_1 = 16$, $\ell_2 = 18$, so $\boldsymbol{\ell_2 - \ell_1 = 2}$.

Communication Complexity of Combinatorial RTS

- The **communication complexity** of an RTS is defined to be the **total number of bits transmitted in the protocol divided by the number of bits in the secret**.
- There are a total of d subshares transmitted to the user whose share is being repaired, where d is the block size of the distribution design.
- The size of the secret is $\ell_2 - \ell_1$ times the size of a subshare.
- Therefore, the communication complexity is

$$\frac{d}{\ell_2 - \ell_1}.$$

- In the projective plane examples from the previous slide, we have $d = 6$. The communication complexity is $\mathbf{6/5}$ when $t = 2$; $\mathbf{3/2}$ when $t = 3$; and $\mathbf{3}$ when $t = 4$.

References

[1] **S. Ball and A. Blokhuis**. **On the size of a double blocking set in $\mathrm{PG}(2, q)$**. *FFA* **2** (1996), 125–137.

[2] **T. M. Laing and D. R. Stinson**. **A survey and refinement of repairable threshold schemes**. *JMC* **12** (2018), 57–81.

[3] **M. Nojoumian**. **Novel Secret Sharing and Commitment Schemes for Cryptographic Applications**. PhD thesis, University of Waterloo, 2012.

[4] **N.B. Shah, K.V. Rashmi and P.V. Kumar. Information-theoretically secure regenerating codes for distributed storage**. *GLOBECOM 2011*, pp. 1–5.

[5] **D.R. Stinson and R. Wei**. **Combinatorial repairability for threshold schemes**. *DCC* **86** (2018), 195–210.

Thank You For Your Attention!