Secret Sharing Revisited

Emilia Käsper

Katholieke Universiteit Leuven
ESAT-COSIC

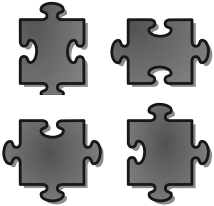
Academy Contact Forum: Coding Theory and Cryptography II
Brussels, September 2007

Introduction

- 1979: Shamir and Blakley propose secret sharing.
- Since then: many new schemes, each of them "better" than Shamir's scheme in some aspect.
- In this talk: threshold secret sharing, alternatives to Shamir's scheme.
- NOT in this talk: generalized access structures, multi-party computation.

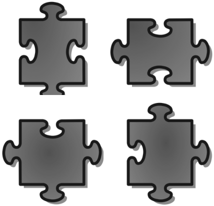
Threshold Secret Sharing

- Secret sharing: n participants hold shares of a secret.



Threshold Secret Sharing

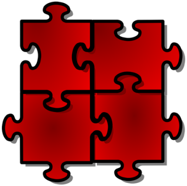
- Secret sharing: n participants hold shares of a secret.



- Perfect secrecy: t participants can learn nothing about the secret.

Threshold Secret Sharing

- Secret sharing: n participants hold shares of a secret.



- Perfect secrecy: t participants can learn nothing about the secret.
- Accessibility: $t + 1$ participants can recover the secret.

Shamir's Secret Sharing

- Take a secret $s \in \mathbb{F}_q$ and t random field elements $a_1, a_2, \ldots, a_t$.
- Define a polynomial

$$f(x) = s + a_1 x + \cdots + a_t x^t \ .$$

- Give participant i, $i = 1, \ldots, n$ point $f(i)$ as a share.
- Lagrange interpolation shows that $t + 1$ participants can recover the polynomial and thus s.
- Given t shares, we can find a polynomial through these points and any secret $s' = f(0)$.
- Thus, t or fewer participants have no information about the secret.

Engineering Aspects of Shamir's Scheme I

- The underlying algebraic structure influences computation efficiency.
- Lagrange interpolation requires that we work over a field.
- In a field $\mathbb{F}_q$, we can have at most $q-1$ participants.
- Thus, we cannot use "natural" structures such as $\mathbb{F}_2$ or $\mathbb{Z}_{32}$.
- Is there a threshold scheme for these structures?

MDS codes

- Let n denote code length, k dimension and d minimum distance.
- Singleton bound for an $[n, k, d]$ code:

$$d \leq n - k + 1$$

- Codes that satisfy the bound with equality are Maximum Distance Separable (MDS) codes.
- Example: Reed-Solomon codes. A message $(a_0, a_1, \ldots, a_{k-1})$ defines a polynomial $f(x) = a_0 + a_1 x + \cdots + a_{k-1} x^{k-1}$. The codeword is

$$(f(1), f(2), \ldots, f(n)) \ .$$

- A $[n, k]$ Reed-Solomon code can correct $n - k$ **erasures**.

Secret Sharing and MDS codes

- Shamir's scheme is a Reed-Solomon code: a secret $f(0)$ is "encoded" as a codeword

$$(f(1), f(2), \dots, f(n)) \ .$$

- Missing shares correspond to erasures in the code.
- An $[n+1, k]$ Reed-Solomon code defines a $(k-1, n)$ threshold scheme.
- In fact, every (t, n) linear threshold secret sharing scheme is equivalent to some $[n+1, t+1]$ MDS code.
- Do there exist MDS codes with $q \leq n$?

Main Conjecture on MDS Codes

If $\mathcal{C}$ is a $[n+1, k, d]$ MDS code over $\mathbb{F}_q$, then

$$\begin{aligned} n &\le k \quad \text{for } q \le k \ , \\ n &\le q+1 \quad \text{for } k = 3 \text{ and } k = q-1 \text{ and } q \text{ even} \ , \\ n &\le q \quad \text{otherwise} \ . \end{aligned}$$

- The first case corresponds to a $(n-1, n)$ scheme.
- In all other cases, the number of participants n is bound by the field size q.

Secret Sharing over Small Fields

- Every linear $[n+1, k, d]$ code $\mathcal{C}$ defines a secret sharing scheme such that
  - $d^{\perp} - 2$ participants learn nothing about the secret;
  - $n - d + 2$ participants can recover the secret.
- Singleton bound implies $d^{\perp} - 2 < n - d + 2$.
- Question: can t participants recover the secret, if

$$d^{\perp} - 2 < t < n - d + 2 \ .$$

- Answer: sometimes.
- We can work over a small field, but we only get a quasi-threshold structure.

Secret Sharing over Small Fields

- Option 1 [CCGHV07]: use a random code.
  - We can work over $\mathbb{F}_2$.
  - Bounds on minimum distance are probabilistic and asymptotic.
- Option 2 [CC06]: use higher order curves.
  - Elliptic curves over $\mathbb{F}_q$ allow up to $q + 2\sqrt{q}$ participants.
  - The case $q = 2$ has no strong bearing.
  - Higher order curves—efficient?

Secret Sharing over Groups

- Can we avoid field arithmetic altogether?
- It would be nice to work over $\mathbb{Z}_{2^k}\dots$
- Shamir's scheme/Lagrange interpolation does not work
- Example over $\mathbb{Z}_{16}$:

$$f(x) = s + a_1x + a_2x^2$$

- $f(1)$, $f(3)$, $f(5)$ together have no information about the secret.
- Individual shares leak information:

$$f(2) = s + 2a_1 + 4a_2 \equiv s \bmod 2$$

Secret Sharing over Groups

- In [CF02]: secret sharing over arbitrary Abelian groups
- Employs a ring of polynomials $\mathcal{S} = \mathbb{Z}[X]/(f(X))$ such that $\deg(f) \approx \log n$
- Each participant gets $\approx \log n$ shares:

$$I \approx \frac{1}{\log n}$$

- For black-box group constructions, the information rate is best possible.
- We can work over groups, but the information rate is sub-optimal.

Engineering Aspects of Shamir's Scheme II

- Traditional use of secret sharing: small secrets (keys).
- Suppose we want to share a **large** secret.
- Share size has impact on computation and communication.
- To share an m-bit secret amongst n players, we need to distribute nm bits.
- To recover a secret, we need to retrieve $(t+1)m$ bits.

Simple Ramp Secret Sharing

- Secrets $s_0, s_1, \ldots, s_{\ell-1}$
- Pick a degree $t + \ell - 1$ polynomial $f(x)$ subject to

$$f(0) = s_0,\ f(1) = s_1, \ldots, f(\ell - 1) = s_{\ell-1}\ .$$

- Give $f(\ell), \ldots, f(n)$ as shares.
- Gradual leakage:
  - t participants have no information.
  - Each additional share leaks $\log q$ bits of information.
  - $t + \ell$ participants recover all secrets.
- This is a $(t, t + \ell, n)$ ramp scheme, where $n \leq q - \ell$.

Trade-Offs in Ramp Secret Sharing

- A $(t, t+\ell, n)$ ramp scheme has information rate $I = \ell$.
- We expand ℓm bits into nm bits in shares.
- We need to retrieve $(t+\ell)m$ bits for recovery—overhead tm.
- Setting $t = 0$ gives optimal information rate.
- We trade secrecy for communication and storage complexity.

Ramp Secret Sharing and Information Dispersal

- A $(0, t, n)$ ramp scheme has optimal information rate.
- A $(0, t, n)$ scheme is simply an information dispersal scheme.
- Each individual share leaks information.
- Option 1: accept gradual leakage.
- Option 2: disperse encrypted data with a $(0, t, n)$ ramp scheme, share key with (t, n) Shamir's scheme.

Conclusions

- Many schemes are better than Shamir's scheme in some parameter.
- … but there is always a trade-off with another parameter.
- A "perfect" scheme does not exist:
  - MDS conjecture bounds number of participants.
  - Black-box schemes over groups cannot have information rate $I = 1$.
- Ramp schemes trade security for communication and storage.
- Ramp schemes optimized for high information rate—information dispersal schemes.