Secret Sharing

A secret sharing scheme is a means for n parties to carry *shares* or *parts* s_i of a message s, called the *secret*, such that the complete set $s_1, \ldots s_n$ of the parts determines the message. The secret sharing scheme is said to be *perfect* if no proper subset of shares leaks any information regarding the secret.

Two party secret sharing. Let s be a secret, encoding as an integer in $\mathbb{Z}/m\mathbb{Z}$. Let $s_1 \in \mathbb{Z}/m\mathbb{Z}$ be generated at random by a trusted party. Then the two shares are defined to be s_1 and $s - s_1$. The secret is recovered as $s = s_1 + s_2$.

Multiple party secret sharing. Let $s \in \mathbb{Z}/m\mathbb{Z}$ be a secret to be shared among n parties. Generate the first $n-1$ shares $s_1, \ldots, s_{n-1}$ at random and set

$$s_n = s - \sum_{i=1}^{n-1}.$$

The secret is recovered as $s = \sum_{i=1}^{n} s_i$.

A (t, n) *threshold* secret sharing scheme is a method for n parties to carry shares s_i of a message s such that any t of the them to reconstruct the message, but so that no $t-1$ of them can easy do so. The threshold scheme is *perfect* if knowledge of $t-1$ or fewer shares provides no information regarding s.

Shamir's (t, n)-threshold scheme. A scheme of Shamir provide an elegant construction of a perfect (t, n)-threshold scheme using a classical algorithm called Lagrange interpolation. First we introduce Lagrange interpolation as a theorem.

Theorem 10 (Lagrange interpolation) *Given t distinct points (x_i, y_i) of the form $(x_i, f(x_i))$, where $f(x)$ is a polynomial of degree less that t, then $f(x)$ is determined by*

$$f(x) = \sum_{i=1}^{t} y_i \prod_{\substack{1 \le j \le t \\ i \ne j}} \frac{x - x_j}{x_i - x_j}. \tag{3}$$

Shamir's scheme is defined for a secret $s \in \mathbb{Z}/p\mathbb{Z}$ with p prime, by setting $a_0 = s$, and choosing $a_1, \ldots, a_{t-1}$ at random in $\mathbb{Z}/p\mathbb{Z}$. The trusted party computes $f(i)$, where

$$f(x) = \sum_{k=0}^{t-1} a_k x^k,$$

for all $1 \le i \le n$. The shares $(i, f(i))$ are distributed to the n distinct parties. Since the secret is the constant term $s = a_0 = f(0)$, the secret is reovered from any t shares $(i, f(i))$, for $I \subset \{1, \ldots, n\}$ by

$$s = \sum_{i \in I} c_i f(i), \text{ where each } c_i = \prod_{\substack{j \in I \\ j \ne i}} \frac{i}{j - i}.$$

Exercise. Verify the correctness of the formula for the secret by substituting into the formula of Lagrange's interpolation theorem.

Properties. Shamir's secret sharing scheme is (1) *perfect* — no information is leaked by the shares, (2) *ideal* — every share is of the same size p as the secret, and (3) involves no unproven hypotheses. In comparison, most public key cryptosystems rely on certain well-known problems (integer factorization, discrete logarithm problems) to be hard in order to guarantee security.

Proof of Lagrange interpolation theorem. Let $g(x)$ be the right hand side of (3). For each x_i in we verify directly that $f(x_i) = g(x_i)$, so that $f(x) - g(x)$ is divisible by $x - x_i$. It follows that

$$\prod_{i=1}^{t}(x - x_i)\big|(f(x) - g(x)), \tag{4}$$

but since $\deg(f(x) - g(x)) \leq t$, the only polynomial of this degree satisfying equation (4) is $f(x) - g(x) = 0$.

Example. Shamir secret sharing with $p = 31$. Let the threshold be $t = 3$, and the secret be $7 \in \mathbb{Z}/31\mathbb{Z}$. We choose elements at random $a_1 = 19$ and $a_2 = 21$ in $\mathbb{Z}/31\mathbb{Z}$, and set $f(x) = 7 + 19x + 21x^2$. As the trusted pary, we can now generate as many shares as we like,

$$\begin{array}{ll} (1, f(1)) = (1, 16) & (5, f(5)) = (5, 7) \\ (2, f(2)) = (2, 5) & (6, f(6)) = (6, 9) \\ (3, f(3)) = (3, 5) & (7, f(7)) = (7, 22) \\ (4, f(4)) = (4, 16) & (8, f(8)) = (8, 15) \end{array}$$

which are distributed to the holders of the share recipients, and the original polynomial $f(x)$ is destroyed. The secret can be recovered from the formula

$$f(x) = \sum_{i=1}^{t} y_i \prod_{\substack{1\leq i\leq t\\ i\neq j}} \frac{x - x_j}{x_i - x_j} \quad =\rangle \quad f(0) = \sum_{i=1}^{t} y_i \prod_{\substack{1\leq i\leq t\\ i\neq j}} \frac{x_j}{x_j - x_i}$$

using any t shares $(x_1, y_1), \ldots, (x_t, y_t)$. If we take the first three shares $(1, 16)$, $(2, 5)$, $(3, 5)$, we compute

$$\begin{aligned} f(0) &= \frac{16 \cdot 2 \cdot 3}{(1-2)(1-3)} + \frac{5 \cdot 1 \cdot 3}{(2-1)(2-3)} + \frac{5 \cdot 1 \cdot 2}{(3-1)(3-2)} \\ &= 3 \cdot 2^{-1} + 15 \cdot (-1) + 10 \cdot 2^{-1} = 17 - 15 + 5 = 7. \end{aligned}$$

This agrees with the same calculation for the shares $(1, 16)$, $(5, 7)$, and $(7, 22)$,

$$\begin{aligned} f(0) &= \frac{16 \cdot 5 \cdot 7}{(1-5)(1-7)} + \frac{7 \cdot 1 \cdot 7}{(5-1)(5-7)} + \frac{22 \cdot 1 \cdot 5}{(7-1)(7-5)} \\ &= 2 \cdot 24^{-1} + 18 \cdot (-8)^{-1} + 17 \cdot 12^{-1} = 13 + 21 + 4 = 7. \end{aligned}$$