Chapter 17 Secret Sharing Schemes

Imagine, if you will, that you have made billions of dollars from Internet stocks and you wish to leave your estate to relatives. Your money is locked up in a safe whose combination only you know. You don’t want to give the combination to each of your seven children because they are less than trustworthy. You would like to divide it among them in such a way that three of them have to get together to reconstruct the real combination. That way, someone who wants some of the inheritance must somehow cooperate with two other children. In this chapter we show how to solve this type of problem.

17.1 Secret Splitting

The first situation that we present is the simplest. Consider the case where you have a message M, represented as an integer, that you would like to split between two people Alice and Bob in such a way that neither of them alone can reconstruct the message M. A solution to this problem readily lends itself: Give Alice a random integer r and give Bob $M - r$. In order to reconstruct the message M, Alice and Bob simply add their pieces together.

A few technical problems arise from the fact that it is impossible to choose a random integer in a way that all integers are equally likely (the sum of the infinitely many equal probabilities, one for each integer, cannot equal 1). Therefore, we choose an integer n larger than all possible messages M that might occur and regard M and r as numbers mod n. Then there is no problem choosing r as a random integer mod n; simply assign each integer mod n the probability $1/n$.

Now let us examine the case where we would like to split the secret among three people, Alice, Bob, and Charles. Using the previous idea, we choose two random numbers r and s mod n and give $M - r - s \pmod{n}$ to Alice, r to Bob, and s to Charles. To reconstruct the message M, Alice, Bob, and Charles simply add their respective numbers.

For the more general case, if we wish to split the secret M among m people, then we must choose $m - 1$ random numbers $r_1, \ \ldots \ , r_{m-1}$ mod n and give them to $m - 1$ of the people, and $M - \sum_{k=1}^{m-1} r_k \pmod{n}$ to the remaining person.

17.2 Threshold Schemes

In the previous section, we showed how to split a secret among m people so that all m were needed in order to reconstruct the secret. In this section we present methods that allow a subset of the people to reconstruct the secret.

It has been reported that the control of nuclear weapons in Russia employed a safety mechanism where two out of three important people were needed in order to launch missiles. This idea is not uncommon. It’s in fact a plot device that is often employed in spy movies. One can imagine a control panel with three slots for keys and the missile launch protocol requiring that two of the three keys be inserted and turned at the same time in order to launch missiles to eradicate the earth.

Why not just use the secret splitting scheme of the previous section? Suppose some country is about to attack the enemy of the week, and the secret is split among three officials. A secret splitting method would need all three in order to reconstruct the key needed for the launch codes. This might not be possible; one of the three might be away on a diplomatic mission making peace with the previous week’s opponent or might simply refuse because of a difference of opinion.

Definition

Let t, w be positive integers with $t \leq w$. A (t, w)-**threshold scheme** is a method of sharing a message M among a set of w participants such that any subset consisting of t participants can reconstruct the message M, but no subset of smaller size can reconstruct M.

The $(t,\, w)$-threshold schemes are key building blocks for more general sharing schemes, some of which will be explored in the Exercises for this chapter. We will describe two methods for constructing a $(t,\, w)$-threshold scheme.

The first method was invented in 1979 by Shamir and is known as the **Shamir threshold scheme** or the Lagrange interpolation scheme. It is based upon some natural extensions of ideas that we learned in high school algebra, namely that two points are needed to determine a line, three points to determine a quadratic, and so on.

Choose a prime p, which must be larger than all possible messages and also larger than the number w of participants. All computations will be carried out mod p. The prime replaces the integer n of Section 17.1. If a composite number were to be used instead, the matrices we obtain might not have inverses.

The message M is represented as a number mod p, and we want to split it among w people in such a way that t of them are needed to reconstruct the message. The first thing we do is randomly select $t-1$ integers mod p; call them $s_1,\, s_2,\, \cdots\, s_{t-1}$. Then the polynomial

$$s(x) \equiv M + s_1 x + \cdots + s_{t-1} x^{t-1} \pmod{p}$$

is a polynomial such that $s(0) \equiv M \pmod{p}$. Now, for the w participants, we select distinct integers $x_1,\, \ldots,\, x_w \pmod{p}$ and give each person a pair $(x_i,\, y_i)$ with $y_i \equiv s(x_i) \pmod{p}$. For example, $1,\, 2,\, \ldots,\, w$ is a reasonable choice for the x’s, so we give out the pairs $(1,\, s(1)),\, \ldots,\, (w,\, s(w))$, one to each person. The prime p is known to all, but the polynomial $s(x)$ is kept secret.

Now suppose t people get together and share their pairs. For simplicity of notation, we assume the pairs are

$(x_1, y_1), \cdots, (x_t, y_t)$. They want to recover the message M.

We begin with a linear system approach. Suppose we have a polynomial $s(x)$ of degree $t-1$ that we would like to reconstruct from the points $(x_1, y_1), \cdots, (x_t, y_t)$, where $y_k = s(x_k)$. This means that

$$y_k \equiv M + s_1 x_k^1 + \cdots + s_{t-1} x_k^{t-1} \pmod{p}, \; 1 \le k \le t.$$

If we denote $s_0 = M$, then we may rewrite this as

$$\begin{matrix} 1 & x_1 & \cdots & x_1^{t-1} \\ 1 & x_2 & \cdots & x_2^{t-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_t & \cdots & x_t^{t-1} \end{matrix} \quad \begin{matrix} s_0 \\ s_1 \\ \vdots \\ s_{t-1} \end{matrix} \quad \equiv \quad \begin{matrix} y_1 \\ y_2 \\ \vdots \\ y_t \end{matrix} \quad \pmod{p}.$$

The matrix, let's call it V, is what is known as a Vandermonde matrix. We know that this system has a unique solution mod p if the determinant of the matrix V is nonzero mod p (see Section 3.8). It can be shown that the determinant is

$$\det V = \prod_{1 \le j < k \le t} (x_k - x_j),$$

which is zero mod p only when two of the x_i's coincide mod p (this is where we need p to be prime; see Exercise 13(a) in Chapter 3). Thus, as long as we have distinct x_k's, the system has a unique solution.

We now describe an alternative approach that leads to a formula for the reconstruction of the polynomial and hence for the secret message. Our goal is to reconstruct a polynomial $s(x)$ given that we know t of its values (x_k, y_k). First, let

$$l_k(x) = \prod_{\substack{i=1 \\ i \ne k}}^{t} \frac{x - x_i}{x_k - x_i} \pmod{p}.$$

Here, we work with fractions mod p as described in Section 3.3. Then

$$l_k(x_j) \equiv \begin{cases} 1 \text{ when } k = j \\ 0 \text{ when } k \neq j. \end{cases}$$

This is because $l_k(x_k)$ is a product of factors $(x_k - x_i)/(x_k - x_i)$, all of which are 1. When $k \neq j$, the product for $l_k(x_j)$ contains the factor $(x_j - x_j)/(x_k - x_j)$, which is 0.

The **Lagrange interpolation polynomial**

$$p(x) = \sum_{k=1}^{t} y_k l_k(x)$$

satisfies the requirement $p(x_j) = y_j$ for $1 \leq j \leq t$. For example,

$$p(x_1) = y_1 l_1(x_1) + y_2 l_2(x_2) + \cdots \equiv y_1 \cdot 1 + y_2 \cdot 0 + \cdots \equiv y_1 \pmod{p}.$$

We know from the previous approach with the Vandermonde matrix that the polynomial $s(x)$ is the only one of degree $t - 1$ that takes on the specified values. Therefore, $p(x) = s(x)$.

Now, to reconstruct the secret message all we have to do is calculate $p(x)$ and evaluate it at $x = 0$. This gives us the formula

$$M \equiv \sum_{k=1}^{t} y_k \prod_{\substack{j=1 \\ j \neq k}}^{t} \frac{-x_j}{x_k - x_j} \pmod{p}.$$

Example

Let’s construct a $(3, 8)$-threshold scheme. We have eight people and we want any three to be able to determine the secret, while two people cannot determine any information about the message.

Suppose the secret is the number $M = 190503180520$ (which corresponds to the word *secret*). Choose a prime p, for example, $p = 1234567890133$ (we need a prime at least as large as the secret, but there is no advantage in using primes much larger than the maximum size of the secret). Choose random numbers s_1 and s_2 mod p and form the polynomial

$$s(x) = M + s_1 x + s_2 x^2.$$

For example, let's work with

$$s(x) = 190503180520 + 482943028839x + 1206749628665x^2.$$

We now give the eight people pairs $(x,\ s(x))$. There is no need to choose the values of x randomly, so we simply use $x = 1,\ 2,\ \dots,\ 8$. Therefore, we distribute the following pairs, one to each person:

$$\begin{array}{l}
(1,\ 645627947891) \\
(2,\ 1045116192326) \\
(3,\ 154400023692) \\
(4,\ 442615222255) \\
(5,\ 675193897882) \\
(6,\ 852136050573) \\
(7,\ 973441680328) \\
(8,\ 1039110787147).
\end{array}$$

Suppose persons 2, 3, and 7 want to collaborate to determine the secret. Let's use the Lagrange interpolating polynomial. They calculate that the following polynomial passes through their three points:

$$20705602144728/5 - 1986192751427x + (1095476582793/5)x^2.$$

At this point they realize that they should have been working mod p. But

$$740740734080 \times 5 \equiv 1 \pmod{p},$$

so they replace 1/5 by 740740734080, as in Section 3.3, and reduce mod p to obtain

$$190503180520 + 482943028839x + 1206749628665x^2.$$

This is, of course, the original polynomial $s(x)$. All they care about is the constant term 190503180520, which is the secret. (The last part of the preceding calculations could have been shortened slightly, since they only needed the constant term, not the whole polynomial.)

Similarly, any three people could reconstruct the polynomial and obtain the secret.

If persons 2, 3, and 7 chose the linear system approach instead, they would need to solve the following:

$$\begin{pmatrix} 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 7 & 49 \end{pmatrix} \begin{pmatrix} M \\ s_1 \\ s_2 \end{pmatrix} \equiv \begin{pmatrix} 1045116192326 \\ 154400023692 \\ 973441680328 \end{pmatrix} \pmod{1234567890133}.$$

This yields

$$(M, s_1, s_2) \equiv (190503180520, 482943028839, 1206749628665),$$

so they again recover the polynomial and the message.

What happens if only two people get together? Do they obtain any information? For example, suppose that person 4 and person 6 share their points (4, 442615222255) and (6, 852136050573) with each other. Let c be any possible secret. There is a unique quadratic polynomial $ax^2 + bx + c$ passing through the points $(0, c)$, (4, 442615222255), and (6, 852136050573). Therefore, any secret can still occur.

Similarly, they cannot guess the share held, for example, by person 7: Any point $(7, y_7)$ yields a unique secret c, and any secret c yields a polynomial $ax^2 + bx + c$, which corresponds to $y_7 = 49a + 7b + c$. Therefore, every value of y_7 can occur, and each corresponds to a secret. So persons 4 and 6 don’t obtain any additional information about which secrets are more likely when they have only their own two points.

Similarly, if we use a polynomial of degree $t - 1$, there is no way that $t - 1$ persons can obtain information about the message with only their data. Therefore, t people are required to obtain the message.

For another example, see Example 38 in the Computer Appendices.

There are other methods that can be used for secret sharing. We now describe one due to Blakley, also from 1979. Suppose there are several people and we want to arrange that any three can find the secret, but no two can. Choose a prime p and let x_0 be the secret. Choose y_0, z_0 randomly mod p. We therefore have a point $Q = (x_0, y_0, z_0)$ in three-dimensional space mod p. Each person is given the equation of a plane passing through Q. This is accomplished as follows. Choose a, b randomly mod p and then set $c \equiv z_0 - ax_0 - by_0 \pmod{p}$. The plane is then

$$z = ax + by + c.$$

This is done for each person. Usually, three planes will intersect in a point, which must be Q. Two planes will intersect in a line, so usually no information can obtained concerning the secret x_0 (but see Exercise 13).

Note that only one coordinate should be used to carry the secret. If the secret had instead been distributed among all three coordinates x_0, y_0, z_0, then there might be only one meaningful message corresponding to a point on a line that is the intersection of two persons' planes.

The three persons who want to deduce the secret can proceed as follows. They have three equations

$$a_i x + b_i y - z \equiv -c_i \pmod{p}, \; 1 \leq i \leq 3,$$

which yield the matrix equation

$$\begin{matrix} a_1 & b_1 & -1 \\ a_2 & b_2 & -1 \\ a_3 & b_3 & -1 \end{matrix} \quad \begin{matrix} x_0 \\ y_0 \\ z_0 \end{matrix} \quad \equiv \quad \begin{matrix} -c_1 \\ -c_2 \\ -c_3 \end{matrix} \; .$$

As long as the determinant of this matrix is nonzero mod p, the matrix can be inverted mod p and the secret x_0 can be found (of course, in practice, one would tend to solve this by row operations rather than by inverting the matrix).

Example

Let $p = 73$. Suppose we give A, B, C, D, E the following planes:

$$\begin{aligned} A: z &= 4x + 19y + 68 \\ B: z &= 52x + 27y + 10 \\ C: z &= 36x + 65y + 18 \\ D: z &= 57x + 12y + 16 \\ E: z &= 34x + 19y + 49. \end{aligned}$$

If A, B, C want to recover the secret, they solve

$$\begin{matrix} 4 & 19 & -1 \\ 52 & 27 & -1 \\ 36 & 65 & -1 \end{matrix} \quad \begin{matrix} x_0 \\ y_0 \\ z_0 \end{matrix} \quad \equiv \quad \begin{matrix} -68 \\ -10 \\ -18 \end{matrix} \pmod{73}.$$

The solution is $(x_0, y_0, z_0) = (42, 29, 57)$, so the secret is $x_0 = 42$. Similarly, any three of A, B, C, D, E can cooperate to recover x_0.

By using $(t-1)$-dimensional hyperplanes in t-dimensional space, we can use the same method to create a (t, w)-threshold scheme for any values of t and w.

As long as p is reasonably large, it is very likely that the matrix is invertible, though this is not guaranteed. It would not be hard to arrange ways to choose a, b, c so that the matrix is always invertible. Essentially, this is what happens in the Shamir method. The matrix equations for both methods are similar, and the Shamir method could be regarded as a special case of the Blakley method. But since the Shamir method yields a Vandermonde matrix, the equations can always be solved. The other advantage of the Shamir method is that it requires less information to be carried by each person: (x, y) versus $(a, b, c, \ \dots)$.

We now return to the Shamir method and consider variations of the basic situation. By giving certain persons more shares, it is possible to make some people more important than others. For example, suppose we have a system in which eight shares are required to obtain the secret, and suppose the boss is given four shares, her daughters are given two shares, and the other employees are each given one share. Then the boss and two of her daughters can obtain the secret, or three daughters and two regular employees, for example.

Here is a more complicated situation. Suppose two companies A and B share a bank vault. They want a system where four employees from A and three from B are needed in order to obtain the secret combination. Clearly it won't work if we simply supply shares that are all for the same secret, since one company could simply use enough partial secrets from its employees that the other company's shares would not be needed. The following is a solution that works. Write the secret s as the sum of two numbers $s \equiv c_A + c_B \pmod{p}$. Now make c_A into a secret shared among the employees of A as the constant term of a polynomial of degree 3. Similarly, let c_B be the constant term of a polynomial of degree 2 and use this to distribute shares of c_B among the employees of B. If four employees of A and three employees of B get together, then those from A determine c_A and those from B determine c_B. Then they add c_A and c_B to get s.

Note that A does not obtain any information about the secret s by itself since $c_A + x \equiv s \pmod{p}$ has a unique solution x for every s, so every possible value of s corresponds to a possible value of c_B. Therefore, knowing c_A does not help A to find the secret; A also needs to know c_B.

17.3 Exercises

1. Suppose you have a secret, namely 5. You want to set up a system where four persons A, B, C, D are given shares of the secret in such a way that any two of them can determine the secret, but no one alone can determine the secret. Describe how this can be done. In particular, list the actual pieces of information (i.e., numbers) that you could give to each person to accomplish this.

2. Persons A, B, C participate in a Shamir $(3, 2)$ secret sharing scheme. They work mod 11. A receives the share $(1, 5)$, B receives $(2, 9)$, and C receives $(3, 3)$.

    1. Show that at least one of the three shares is incorrect.

    2. Suppose A and C have correct shares. Find the secret.

3. You set up a (2, 30) Shamir threshold scheme, working mod the prime 101. Two of the shares are (1,13) and (3,12). Another person received the share (2, *), but the part denoted by * is unreadable. What is the correct value of * ?

4. You set up a (2, 10) Shamir threshold scheme, working mod the prime 73. Two of the shares are (1, 10) and (2, 18). A third share is (5, *). What is *?

5. In a $(3, 5)$ Shamir secret sharing scheme with modulus $p = 17$, the following were given to Alice, Bob, and Charles: $(1, 8)$, $(3, 10)$, $(5, 11)$. Calculate the corresponding Lagrange interpolating polynomial, and identify the secret.

6. In a Shamir secret sharing scheme, the secret s is the constant term of a degree 4 polynomial mod the prime 1093. Suppose three people have the secrets (2, 197), (4, 874), and (13, 547). How many possibilities are there for the secret? (*Note:* We assume that $0 \leq s \leq 1092$.)

7. Mark doesn't like mods, so he wants to implement a $(2, 30)$ Shamir secret sharing scheme without them. His secret is M (a positive integer) and he gives person i the share $(i, M + si)$ for a positive integer s that he randomly chooses. Bob receives the share $(20, 97)$. Describe how Bob can narrow down the possibilities for M and determine what values of M are possible.

8. A key distributor uses a $(2, 20)$-threshold scheme to distribute a combination to an electronic safe to 20 participants.

1. What is the smallest number of participants needed to open the safe, given that one unknown participant is a cheater who will reveal a random share?

2. If they are only allowed to try one combination (if they are wrong the electronic safe shuts down permanently), then how many participants are necessary to open the safe? (Note: This one is a little subtle. A majority vote actually works with four people, but you need to show that a tie cannot occur.)

9. A certain military office consists of one general, two colonels, and five desk clerks. They have control of a powerful missile but don't want the missile launched unless the general decides to launch it, or the two colonels decide to launch it, or the five desk clerks decide to launch it, or one colonel and three desk clerks decide to launch it. Describe how you would do this with a secret sharing scheme. (Hint: Try distributing the shares of a (10, 30) Shamir scheme.)

10. Suppose there are four people in a room, exactly one of whom is a foreign agent. The other three people have been given pairs corresponding to a Shamir secret sharing scheme in which any two people can determine the secret. The foreign agent has randomly chosen a pair. The people and pairs are as follows. All the numbers are mod 11.

$$A\colon (1, 4) \quad B\colon (3, 7) \quad C\colon (5, 1) \quad D\colon (7, 2).$$

Determine who the foreign agent is and what the message is.

11. Consider the following situation: Government A, Government B, and Government C are hostile to each other, but the common threat of Antarctica looms over them. They each send a delegation consisting of 10 members to an international summit to consider the threat that Antarctica's penguins pose to world security. They decide to keep a watchful eye on their tuxedoed rivals. However, they also decide that if the birds get too rowdy, then they will launch a full-force attack on Antarctica. Using secret sharing techniques, describe how they can arrange to share the launch codes so that it is necessary that three members from delegation A, four members from delegation B, and two members from C cooperate to reconstruct the launch codes.

12. This problem explores what is known as the Newton form of the interpolant. In the Shamir method, we presented two methods for calculating the interpolating polynomial. The system of equations approach is difficult to solve and easy to evaluate, while with the Lagrange approach it is quite simple to determine the interpolating polynomial but becomes a labor to evaluate. The Newton form of the interpolating polynomial comes from choosing
$1, x - x_1, (x - x_1)(x - x_2), \ \cdots \ , (x - x_1)(x - x_2) \ \cdots \ (x - x_t)$
as a basis. The interpolating polynomial is then

$p(x) = c_0 + c_1(x - x_1) + c_2(x - x_1)(x - x_2) + \cdots + c_t(x - x_1)(x - x_2)\ \cdots\ (x - x_t)$. Show that we can solve for the coefficients c_k by solving a system $Nc = y$. What special properties do you observe in the matrix N? Why does this make the system easier to solve?

13. In a Blakley $(3, w)$ scheme, suppose persons A and B are given the planes $z = 2x + 3y + 13$ and $z = 5x + 3y + 1$. Show that they can recover the secret without a third person.

17.4 Computer Problems

1. Alice, Bob, and Charles have each received shares of a secret that was split using the secret splitting scheme described in Section 17.1. Suppose that $n = 2110763$. Alice is given the share $M - r - s = 1008369$, Bob is given the share $r = 593647$, and Charles is given the share $s = 631870$. Determine the secret M.

2. For a Shamir (4,7) secret sharing scheme, take $p = 8737$ and let the shares be

   $$(1,\ 214),\ (2,\ 7543),\ (3,\ 6912),\ (4,\ 8223),\ (5,\ 3904),\ (6,\ 3857),\ (7,\ 510).$$

   Take a set of four shares and find the secret. Now take another set of four shares and verify that the secret obtained is the same.

3. Alice, Bob, Charles, and Dorothy use a (2, 4) Shamir secret sharing scheme using the prime $p = 984583$. Suppose that Alice gets the share (38, 358910), Bob gets the share (3876, 9612), Charles gets the share (23112, 28774), and Dorothy gets the share (432, 178067). One of these shares was incorrectly received. Determine which one is incorrect, and find the secret.