The Knapsack Cryptosystem

Ryan C. Daileda

Trinity University

Number Theory

Introduction

Today we will briefly consider a public key cryptosystem whose security rests on the difficulty in solving a certain combinatorial problem.

The idea is to encode messages using a scrambled "base" system in such a way that only the intended recipient can retrieve the message's "digits" (bits).

Although no longer utilized in practice, this cryptosystem is nonetheless another interesting example.

Knapsack Problems

Suppose we are given a "knapsack" of volume $V \in \mathbb{N}$, and "objects" of volume $a_1, a_2, \ldots, a_r \in \mathbb{N}$.

The *knapsack problem* asks whether or not it it possible to choose a subset of the a_i whose total volume is V.

That is, do there exist $\epsilon_i \in \{0, 1\}$ so that

$$V = \epsilon_1 a_1 + \epsilon_2 a_2 + \cdots + \epsilon_r a_r?$$

In general, this is an extremely difficult question to answer.

Example

Suppose that $a_1 = 2$, $a_2 = 5$, $a_3 = 6$, $a_4 = 9$, $a_5 = 13$.

If $V = 20$, the knapsack problem has a solution, since

$$20 = 0 \cdot 2 + 1 \cdot 5 + 1 \cdot 6 + 1 \cdot 9 + 0 \cdot 13.$$

However, if $V = 23$, there is no solution to the knapsack problem.

To see this, notice that the sum of the first 4 terms is

$$2 + 5 + 6 + 9 = 22,$$

which means that if there is a solution, then $\epsilon_5 = 1$.

But then we must be able to choose from $2, 5, 6$ and 9 and get a sum of 10, which is clearly impossible.

Superincreasing Sequences

Definition

We say that a sequence $\{a_i\}$ is *superincreasing* if

$$a_n > a_{n-1} + a_{n-2} + \cdots + a_1$$

for all $n \geq 2$.

Examples.

- The sequence $a_1 = 3,\ a_2 = 4,\ a_3 = 10,\ a_4 = 20$ is superincreasing.
- The sequence $a_1 = 1,\ a_2 = 2,\ a_3 = 4,\ a_4 = 8,\ a_5 = 16$ is superincreasing.
- More generally, given $b \geq 2$, the sequence $a_1 = 1,\ a_2 = b,$ $a_3 = b^2, \ldots, a_i = b^{i-1}, \ldots, a_r = b^{r-1}$ is superincreasing.

An important feature of solutions to knapsack problems involving superincreasing sequences is that they are unique.

Theorem 1

Let $\{a_i\} \subset \mathbb{N}$ be a superincreasing sequence. If $\epsilon_i, \delta_i \in \{0, 1\}$ and

$$\sum_{i=1}^{r} \epsilon_i a_i = \sum_{i=1}^{r} \delta_i a_i,$$

then $\epsilon_i = \delta_i$ for all i.

Proof. We induct on r. If $r = 1$, the result is trivial, since $\epsilon_1 a_1 = \delta_1 a_1$ implies $\epsilon_1 = \delta_1$, as $a_1 \neq 0$.

Now let $r > 1$ and assume the result for all superincreasing sequences of length $< r$.

Then

$$\sum_{i \le r} \epsilon_i a_i = \sum_{i \le r} \delta_i a_i \;\Rightarrow\; |\epsilon_r - \delta_r|\, a_r = \left| \sum_{i<r} (\delta_r - \epsilon_r) a_i \right| \le \sum_{i<r} a_i < a_r.$$

Since $|\epsilon_r - \delta_r| \le 1$, this implies that $\epsilon_r = \delta_r$. Thus

$$\sum_{i<r} \epsilon_i a_i = \sum_{i<r} \delta_i a_i,$$

and $\epsilon_i = \delta_i$ for $i < r$ by the inductive hypothesis.

Therefore $\epsilon_i = \delta_i$ for all $i \le r$. This completes the inductive step, and the proof. □

It is relatively easy to solve a knapsack problem involving a superincreasing sequence $\{a_i\}$.

Suppose we are given $V \in \mathbb{N}$ and we want to determine $\epsilon_i \in \{0, 1\}$ so that

$$V = \sum_{i=1}^{r} \epsilon_i a_i.$$

Let n be the largest index so that $a_n \le V$. Then $a_i > V$, and hence $\epsilon_i = 0$, for $i > n$.

Furthermore,

$$\sum_{i<n} a_i < a_n \le V,$$

which means that we can't have $\epsilon_n = 0$. Thus $\epsilon_n = 1$.

Now recursively repeat this procedure for $V - a_n$.

Example

Consider the superincreasing sequence $a_1 = 3$, $a_2 = 4$, $a_3 = 10$, $a_4 = 20$, $a_5 = 42$.

To solve the knapsack problem

$$55 = 3\epsilon_1 + 4\epsilon_2 + 10\epsilon_3 + 20\epsilon_4 + 42\epsilon_5,$$

we start by observing $42 < 55$, so that we must have $\epsilon_5 = 1$.

We then consider $55 - 42 = 7$. Now we have $a_2 < 7 < a_3$, so $\epsilon_4 = \epsilon_3 = 0$ and $\epsilon_2 = 1$.

Finally we have $a_1 = 3 = 7 - 4$, so that $\epsilon_1 = 1$. Therefore

$$55 = 3 \cdot 1 + 4 \cdot 1 + 10 \cdot 0 + 20 \cdot 0 + 42 \cdot 1.$$

Knapsack Encryption

A public key cryptosystem utilizing the knapsack problem was developed by Merkle and Hellman in 1978.

Every user:

1. Chooses a superincreasing sequence $\{a_i\}_{i=1}^r$, a modulus $m > 2a_r$, and a multiplier a with $(a, m) = 1$.
2. Computes $b_i \equiv aa_i \pmod{m}$.
3. Publishes the encryption key $K_E = \{b_i\}$.
4. Keeps $\{a_i\}_{i=1}^r$, m and a secret.

To encrypt a message to the individual with public key $K_E = \{b_i\}_{i=1}^r$, the sender first converts the plaintext into binary blocks of length r.

A given binary block $P = \epsilon_1\epsilon_2\cdots\epsilon_r$ is converted to the ciphertext block

$$C = \sum_{i=1}^{r} \epsilon_i b_i.$$

Because the transformed sequence $\{b_i\}$ is no longer superincreasing, determining the ϵ_i from C is "hard" for an eavesdropper, even given the sequence $\{b_i\}$.

Unpacking the Knapsack

The decryption key is $K_D = (\{a_i\}, m, b)$, where b satisfies $ab \equiv 1 \pmod{m}$, which is easily computed from a and m using the EA.

The message recipient computes $S \equiv bC \pmod{m}$, with $0 \leq S < m$.

Notice that

$$S \equiv bC \equiv \sum_{i=1}^{r} \epsilon_i b b_i \equiv \sum_{i=1}^{r} \epsilon_i b a a_i \equiv \sum_{i=1}^{r} \epsilon_i a_i \pmod{m}.$$

Because $\{a_i\}$ is superincreasing,

$$0 \leq \sum_{i=1}^{r} \epsilon_i a_i < a_r + a_r = 2a_r < m.$$

It follows that $S = \sum_{i=1}^{r} \epsilon_i a_i$, and the recipient can now compute the plaintext $P = \epsilon_1 \epsilon_2 \cdots \epsilon_r$ using the procedure described earlier.

Example 1

Encrypt the plaintext block 01001 using the superincreasing sequence $\{3, 4, 10, 20, 42\}$ with modulus $m = 90$ and multiplier $a = 17$.

Solution. We first multiply our sequence by 17 (modulo 90), obtaining $\{51, 68, 80, 70, 84\}$.

Our ciphertext is then $C = 1 \cdot 68 + 1 \cdot 84 = \boxed{152}$.

To decrypt we compute $17^{-1} \equiv 53 \pmod{90}$ then multiply:

$$53 \cdot C = 53 \cdot 152 \equiv 62 \cdot 53 \equiv 46 \pmod{90}$$

Since $46 = 1 \cdot 4 + 1 \cdot 42$, we recover the plaintext 01001. □

Remark. While certainly interesting, the Merkle-Hellman knapsack cryptosystem (and its variants) were proven to be insecure during the 1980s.

It turns out that the transformation $b_i \equiv aa_i \pmod{m}$ doesn't sufficiently "disguise" the superincreasing nature of $\{a_i\}$.