Subset Sum Problem

Given a positive integer T and the following sequence of positive integers

$$s_1, s_2, \ldots, s_n,$$

find a sequence $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ such that

$$T = s_{i_1} + s_{i_2} + \ldots + s_{i_k}.$$

Example:

$$\{14, 28, 56, 82, 90, 132, 197, 284, 341\}$$

$$\begin{aligned} 515 &= 341 + 132 + 28 + 14 \\ &= 341 + 90 + 56 + 28 \\ &= 197 + 132 + 90 + 82 + 14 \end{aligned}$$

On the otherhand, there are no subset sums for 516.

This problem has been shown to be "**NP-complete**", which means that (among other things) there is no known polynomial-time algorithm that solves it.

Superincreasing Sequences

Definition: A sequence is said to be *superincreasing* if for all j from 2 to n, one has

$$s_j > \sum_{i=1}^{j-1} s_i.$$

To see if T can be expressed as a subset sum, proceed as follows.

1. If $T < s_1$ or $T > s_1 + s_2 + \cdots + s_n$ then no such subset exists, otherwise
2. Find the largest k such that $s_k \leq T$ and set
$$T = r + s_k.$$
3. Repeat process on r using only $s_1, s_2, \ldots, s_{k-1}$.

Example Write 55 as the sum of distinct powers of 2.

$$\begin{aligned} 55 &= 32 + 23 \\ &= 32 + 16 + 7 \\ &= 32 + 16 + 4 + 3 \\ &= 32 + 16 + 4 + 2 + 1 \end{aligned}$$

Random Superincreasing Sequence

Fix $n \geq 1$ and $k > 1$. Then

1. Let s_1 be a random number between 1 and k.
2. For i from 2 to n, let

$$s_i = s_1 + s_2 + \cdots + s_{i-1} + m_i,$$

   where m_i is a random number between 1 and k

Merkle-Hellman Knapsack Cryptosystem

1. Choose a superincreasing sequence
$$s_1, s_2, \ldots, s_n.$$
2. Choose p to be a large prime such that
$$p > s_1 + s_2 + \cdots + s_n.$$
3. Let a be a random number between 1 and $p - 1$ and publicly announce
$$t_i := as_i \mod p.$$

Encryption Process: To encode a message $(x_1, x_2, \ldots, x_n)$ (made of bits of 0 and 1), one sends the single number
$$C := \sum_{i=1}^{n} x_i t_i.$$

Encryption Process: To decode, we need only solve the subset sum problem for
$$M := a^{-1}C \mod p.$$

Merkle-Hellman: Example

Let

$$\{3, 5, 12, 21, 43\} \quad p = 89 \quad a = 15$$

Therefore, the T sequence is given by:

$$\{45, 75, 2, 48, 22\}$$

Encode 01101 by:

$$\begin{aligned} C &= 0 \cdot 45 + 1 \cdot 75 + 1 \cdot 2 + 0 \cdot 48 + 1 \cdot 22 \\ &= 10 \mod 89 \end{aligned}$$

To decode, since $a^{-1} = 6 \mod 89$, we have that

$$\begin{aligned} M = a^{-1}C = 60 &= 17 + 43 \\ &= 5 + 12 + 43 \\ &= 0 \cdot 3 + 1 \cdot 5 + 1 \cdot 12 + 0 \cdot 21 + 1 \cdot 43 \end{aligned}$$