The Knapsack as a Nonlinear Function

Rainer A. Rueppel and James L. Massey
Institute of Telecommunications
Swiss Federal Institute of Technology
CH-8092 Zurich

8 November 1984

ABSTRACT

The general 0/1 knapsack of order N is specified by N positive integers (or "weights") \(w_1, w_2, \ldots, w_N\) and defines an integer-valued function \(S = x_1 w_1 + x_2 w_2 + \ldots + x_N w_N\), where \(x_i \in \{0,1\}\). Equivalently, the bit \(s_i\) of the radix-two form \([s_{N-1}, \ldots, s_1, s_0]\) of \(S\) is thereby specified as a GF(2)-valued function \(f_i\) over the vector space of N-tuples \(x\) over GF(2). When \(w_1 w_2 = \ldots = w_N = 1\) [i.e., when \(s_i\) for \(i > 0\) is just the "carry bit" \(i\) positions forward when \(x_1, x_2, \ldots, x_N\) are summed], a theorem of Lucas is invoked to show that \(s_i\) when written in algebraic normal form [i.e., as a GF(2) sum of products of the variables] contains all and only the product terms of order \(2^i\). It follows by recursion that in the general case the nonlinear order of \(f_i\) [i.e., the order of the maximum product in its algebraic normal form] is at most \(\min(2^i, N)\) -- and intuitive arguments are given to show that this bound is typically quite tight.

A running key generator for a stream cipher system is proposed in which the output bit is the value of a knapsack function \(f_i\) applied to the state \(x\) of a maximal-length linear feedback shift-register. Results of experiments with pseudorandomly chosen weights show that, with high probability, the linear complexity of the output sequence is equal or close to the maximum attainable with any nonlinear output function \(f\) of nonlinear order \(\min(2^i, N)\) satisfying \(f(0) = 0\).
SUMMARY

The general 0/1 knapsack of order N is specified by N positive integers (or "weights") \( w_1, w_2, \ldots, w_N \) and defines the following integer-valued function

\[
S = \sum_{i=1}^{N} x_i w_i
\]  

(1)

whose domain is the set of all \( N \)-tuples \( x = [x_1, x_2, \ldots, x_N] \) whose components are each either the integer 0 or the integer 1. The mapping (1) from \( x \) to \( S \) is, of course, linear if one extends its domain to the vector space of \( N \)-tuples with rational components.

Let \( [s_M, \ldots, s_1, s_0] \) be the radix-two representation of \( S \) where

\[
M = \left\lfloor \log_2 \sum_{i=1}^{N} w_i \right\rfloor.
\]

Then the knapsack (1) can be viewed as defining \( M + 1 \) GF(2)-valued functions

\[
s_i = f_i(x_1, x_2, \ldots, x_N)
\]  

(2)

on the vector space of \( N \)-tuples \( x = [x_1, x_2, \ldots, x_N] \) over GF(2). Note that \( x_i \) is treated as an integer in (1) but as an element of GF(2) in (2). The aim of this paper is to characterize these functions \( f_i \) in a manner useful for cryptography.

The function \( f_i \) is said to be in algebraic normal form (ANF) when it is expressed as a GF(2) sum of product of its variables (plus possibly a constant 1). For instance,
\[ f_1(x_1, x_2, x_3, x_4) = x_1 + x_1 x_3 + x_2 x_4 + x_2 x_3 x_4 \]

is in ANF and has one linear term, two second-order terms and one third-order term. The **nonlinear order** of \( f_1 \) is the order of the maximum order term in its ANF and is \(-\infty\) if \( f_1 = 0 \).

It will be shown with the aid of a theorem of Lucas that when \( w_1 = w_2 = \ldots = w_N \) [so that \( S \) is just the integer sum of the variables \( x_1, x_2, \ldots, x_N \)] then \( s_i \) is the GF(2) sum of all and only the product terms of order \( 2^i \). [Note that in this case, \( s_i \) for \( i > 0 \) is the "carry bit" to \( i \) positions forward from the least significant bit in the sum]. This result allows one to write \( f_1 \) in a recursive manner for arbitrary weights and shows that the nonlinear order \( K \) of \( f_1 \) is bounded as

\[ K \leq \min(2^i, N). \]  

(3)

Intuitive arguments will be given to show that this bound is typically quite tight and that the ANF of \( f_1 \) typically contains many product terms of each positive order less than \( K \).

As a cryptographic application of the nonlinear knapsack function, a stream cipher is proposed in which the running key generator is a maximal-length linear feedback shift-register, whose state \( x \) is the input to the knapsack function \( f_1 \) that emits the running key bit. The results of experiments with pseudo-randomly chosen knapsack weights for such running key generators will be reported, showing that the linear complexity \( L \) of the output sequence is closely given by

\[ L = \sum_{i=1}^{K^*} \binom{N}{i} \]  

(4)
where $K^*$ denotes the upperbound given in (3). It is well-known that $L$ as given by (4) is the maximum linear complexity of a sequence obtained by applying an output function $f$ of nonlinear order $K^*$ satisfying $f(0) = 0$ to the state of an $N$-stage linear feedback shift-register.