Diss. ETH No. 7714

New Approaches to Stream Ciphers

A dissertation submitted to the
SWISS FEDERAL INSTITUTE
OF TECHNOLOGY ZURICH
for the degree of
Doctor of Science

Submitted by
Rainer A. Rueppel
dipl. El.Ing. ETH
born April 21, 1955
citizen of Schaffhausen

accepted on the recommendation of

Prof. Dr. J.L. Massey, referee
Prof. Dr. G.S. Moschytz, co-referee

1984

ADAG Administration & Druck AG

Abstract

Structures that can be used as key stream generators in stream cipher systems are analyzed and characterized and some new promising structures are proposed.

Linear complexity (which is the length of the shortest linear feedback shift register (LFSR) that can generate a given sequence) is shown to be a useful measure of the unpredictability, or equivalently, the randomness of finite sequences. A characterization of truly random binary sequences of length n is obtained in terms of the expectation and the variance of the associated linear complexity. The variance is shown to equal 86/81 plus an exponentially fast decreasing quantity. The growth of linear complexity with increasing sequence length n is described using a random walk. "Typical" random sequences are shown to have a "typical" linear complexity profile. For the practically interesting case of repeating a finite truly random sequence of length 2^m or 2^m-1, it is shown that the expected linear complexity is close to the period length.

Nonlinear combinations of periodic sequences are analyzed in terms of the linear complexity of the produced output sequence. The algebraic normal form of boolean functions is chosen as reference. In the case of nonlinear operations on distinct phases of a maximal-length (m-)sequence, a matrix approach is given which allows complete analyzability, but is in general computational infeasible for cases of practical interest. Key has derived an upperbound on the linear complexity of the produced sequence which depends on the nonlinear order of the employed function and on the order of the recursion of the m-sequence. The probability of selecting a function whose associated output sequence exhibits a linear complexity substantially smaller than the upperbound is shown to tend to zero with increasing prime order of the recursion of the m-sequence. A large class of functions is exhibited for which the linear complexity of the produced sequence is lowerbound by $\binom{L}{k}$, where L denotes the prime order of the recursion of the m-sequence and where k denotes the nonlinear order of the function. In the case of nonlinear operations on sequences with distinct (but not necessarily irreducible) minimal polynomials, it is shown that the linear complexity of the produced output sequence is equal to the

nonlinear combining function evaluated over the reals with the sequence arguments replaced by the associated linear complexities, provided the roots of each minimal polynomial are simple and lie in an extension field whose degree is relatively prime to the degree of the extension field containing each root of the other minimal polynomials.

The effects of clocking LFSRs at rates higher than the system clock are discussed. It is shown that such multiple clocking results in simulating an LFSR different from the one physically implemented. The simulated LFSR is completely determined by the original LFSR and the associated speed factor. A random sequence generator, suggested by a linear cipher problem, that employs multiple clocking is analyzed in detail.

The general 0/1 knapsack defines an integer-valued function over the vectorspace of binary N-tuples. It is shown how integer addition, when both the integers and their sum are represented in radix-2 form, may be described in GF(2). This leads to a complete GF(2)-description of the knapsack where each bit of the sum is computed by an individual nonlinear function defined over the vector space of binary N-tuples. The nonlinear order of the function computing sum bit i is shown to be at most $\min(2^i,N)$, which bound typically is quite tight.

A running key generator for a stream cipher system is proposed in which a knapsack is applied to the state of a maximal-length LFSR. Results of experiments with randomly chosen weights show that, with high probability, the linear complexity of the sequence defined by the ith sum bit of the knapsack is equal or close to the maximum attainable with any nonlinear output function of order $\min(2^i,N)$.

The nonlinear combining function, when augmented with memory, defines a finite state machine by itself. A large class of such combining functions is derived which exhibit maximum immunity against correlation, a threat to which many recently proposed key stream generators succumb. Real addition of r-ary sequences is shown to define such a correlation resistent structure with memory. When two m-sequences of relatively prime recursion order are added over the reals the linear complexity of the resulting sum sequence is shown in general to be equal or very close to the product of the periods of the input sequences.

Kurzfassung

Strukturen für die Anwendung als Schlüsselstromgeneratoren (key stream generators) in Stromchiffriersystemen (stream ciphers) werden analysiert und charakterisiert; neue Strukturen mit interessanten Eigenschaften werden vorgeschlagen.

Es wird gezeigt, dass die lineare Komplexität (die Länge des kürzesten linear rückgekoppelten Schieberegisters, welches eine gegebene Sequenz erzeugen kann) nützliche Eigenschaften als Mass für die Zufälligkeit endlicher Sequenzen hat. Binäre Zufallssequenzen der Länge n werden mit Hilfe des Erwartungswertes und der Varianz der zugehörigen linearen Komplexität charakterisiert. Die ermittelte Varianz hat den Wert 86/81 plus eine exponentiell schnell verschwindende Grösse. Das Wachstumsverhalten der linearen Komplexität in Funktion der Sequenzlänge wird beschrieben mit Hilfe eines Random-Walk-Arguments. Es wird gezeigt, dass "typische" Zufallssequenzen ein "typisches" Profil der zugehörigen linearen Komplexität besitzen. Für den praktisch interessanten Fall der periodischen Wiederholung endlicher Zufallssequenzen der Länge 2^m oder 2^m-1 wird gezeigt, dass der Erwartungswert der linearen Komplexität nahe bei der Periodenlänge liegt.

Nichtlineare Kombinationen periodischer Sequenzen werden analysiert bezüglich der linearen Komplexität der erzeugten Ausgangssequenz. Als Referenz dient die algebraische Normalform boolescher Funktionen. Für den Fall nichtlinearer Kombinationen von unterschiedlichen Phasen einer maximal-langen (m-)Sequenz wird ein Matrix-Verfahren beschrieben, welches unbedingte Analysierbarkeit erlaubt, aber praktisch im allgemeinen nicht durchführbar ist. Key hat eine obere Schranke für die lineare Komplexität der erzeugten Ausgangssequenz hergeleitet, die von der nichtlinearen Ordnung der verwendeten Funktionen und von der Ordnung der Rekursion der m-Sequenz abhängt. Es wird gezeigt, dass die Wahrscheinlichkeit der Auswahl einer Funktion, deren zugehörige Ausgangssequenz eine lineare Komplexität zeigt, die wesentlich kleiner ist als die obere Schranke, gegen Null geht mit zunehmender Primzahlordnung der Rekursion der m-Sequenz. Eine grosse Klasse von Funktionen wird hergeleitet, für welche die lineare Komplexität der erzeugten Ausgangssequenz durch eine untere Schranke von $\binom{L}{k}$ be-

grenzt ist, wobei L die Primzahlordnung der Rekursion der m-Sequenz und k die nichtlineare Ordnung der Funktion bezeichnet.

Für den Fall nichtlinearer Kombinationen von Sequenzen mit unterschiedlichen (aber nicht notwendigerweise irreduziblen) Minimalpolynomen wird gezeigt, dass die lineare Komplexität der erzeugten Ausgangssequenz sich einfach berechnen lässt als reeller Wert der nichtlinearen Kombinationsfunktion, ausgewertet an den linearen Komplexitäten der entsprechenden Eingangssequenzen, falls die Wurzeln jedes Minimalpolynoms einfach sind und in einem Erweiterungskörper liegen, dessen Grad relativ prim zum Grad des Erweiterungskörpers ist, in dem die Wurzeln der restlichen Minimalpolynome liegen.

Wenn linear rückgekoppelte Schieberegister mit höheren Clockraten getaktet werden als das umgebende System, entstehen interessante Effekte. Es wird gezeigt, dass solches Mehrfachtakten der Simulation eines anderen linear rückgekoppelten Schieberegisters entspricht. Das simulierte Schieberegister ist eindeutig bestimmt durch das ursprüngliche Schieberegister und den gewählten Taktgeschwindigkeitsfaktor. Ein Zufallssequenzgenerator, dessen Struktur durch ein lineares Verschlüsselungsproblem motiviert ist und der solches Mehrfachtakten verwendet, wird analysiert.

Der allgemeine 0/1 Knapsack (Rucksack) definiert eine ganzzahlige Funktion über den Vektorraum der binären N-Tupel. Es wird gezeigt, wie ganzzahlige Addition in GF(2) beschrieben werden kann, falls die ganzen Zahlen und ihre Summe in binärer Form gegeben sind. Darauf aufbauend lässt sich eine komplette GF(2)-Beschreibung des Knapsacks herleiten, in der jedes Bit der Knapsacksumme durch eine individuelle nichtlineare GF(2)-wertige Funktion über den Vektorraum der binären N-Tupel berechnet wird. Die nichtlineare Ordnung der Funktion, welche das Summenbit i berechnet, ist höchstens $\min(2^i, N)$ - eine Schranke, die typischerweise sehr eng ist.

Ein Schlüsselstromgenerator wird vorgeschlagen, in welchem der Zustand eines maximal-langen linear rückgekoppelten Schieberegisters den Eingangswert für einen Knapsack liefert. Simulationsresultate mit zufällig

gewählten Knapsack-Gewichten zeigen, dass die lineare Komplexität der i-ten Summenbit-Sequenz des Knapsacks mit hoher Wahrscheinlichkeit gleich oder nahe der mit einer beliebigen nichtlinearen Funktion der Ordnung $min(2^i,N)$ maximal erreichbaren Komplexität ist.

Wenn die nichtlineare Kombinationsfunktion mit Gedächtnis angereichert wird, dann definiert sie einen eigenen endlichen Automaten. Es wird eine grosse Klasse von Kombinationsfunktionen mit Gedächtnis vorgestellt, welche maximale Immunität gegen Korrelationsattacken ausweisen. Es wird gezeigt, dass die reelle Addition von Sequenzen einer solchen korrelationsresistenen Struktur mit Gedächtnis entspricht. Wenn zwei m-Sequenzen, deren Rekursionen relativ prime Ordnung haben, über die reellen Zahlen addiert werden, dann hat die resultierende Summensequenz eine lineare Komplexität gleich oder nahe der eigenen Periodenlänge (welche das Produkt der Perioden der beiden Eingangssequenzen ist).