PUBLICATIONS DE L'INSTITUT MATHÉMATIQUE
Nouvelle série, tome 64 (78), 1998, 183–204

RECENT ADVANCES IN STREAM CIPHER CRYPTANALYSIS

Jovan Đ. Golić

Communicated by Žarko Mijajlović

Abstract. A survey of recent contributions to the analysis of stream ciphers is presented. Emphasis is put on novel approaches, concepts, and results. Related open mathematical and research problems are also pointed out.

Introduction. Over the centuries, cryptography has been developed and used as a tool to protect secret information, especially within the military, diplomatic, and government communities in general. Public interest in cryptography has dramatically increased with the invention of so-called public-key cryptography [**7**] in 1976 which made it possible to achieve information privacy without prior exchange of a secret key over a secure communication channel.

On the other hand, in the modern information era when information is processed, transmitted, and stored in electronic form in large communication and computer networks, this information becomes very vulnerable to being read, copied, and altered without authorization. Cryptographic techniques provide a means to ensure information privacy and authenticity as well as many other related information security objectives. Accordingly, over the years, commercial and civil applications of cryptography has spread from classical communication systems like cable telephony, telegraphy, fascimil, television, and wireless communications to mobile wireless telephony and various services on integrated communication networks (like Internet) as well as to electronic medical files, electronic data interchange, and electronic commerce and banking including smart-cards and electronic money. To ensure information security, cryptographic tools have to be used together with more general (and less mathematical) computer and information security methods and techniques.

Information security provided can be either theoretical (unconditional) or practical (conditional) depending on whether the assumed opponent has unlimited or limited computational resources, respectively. Cryptographic techniques are

AMS Subject Classification (1991): Primary 94A60

essentially mathematical so that cryptology, as the science of cryptography and cryptanalysis, can be regarded as a specific branch of mathematics, computer science, and electrical engineering which incorporates a number of fields including algebra, number theory, probability and statistics, coding and information theory, combinatorics, discrete optimization, algorithms and complexity theory, and communications theory. It has created many interesting and difficult mathematical problems and has given rise as such to a number of scientific fields including information, communications, and complexity theories. Apart from contributing to the creation and progress of theoretical sciences, it has also provided a major incentive for the invention and continuous development of computer technologies.

Stream ciphers are the most practical tool for the so-called private-key (or symmetric-key) encryption, where the communicating parties share the same secret key. In a stream cipher, a sequence of plaintext symbols is transformed into a sequence of ciphertext symbols by a simple encryption transformation applied to plaintext symbols which is chosen according to a sequence of secret key symbols called the keystream sequence. In his seminal paper [**77**], Shannon has shown that the *perfect secrecy* (in the information-theoretic, unconditional sense) can be achieved by using a *purely random* keystream sequence, that is, a sequence of independent uniformly distributed random variables. If the keystream sequence is generated pseudorandomly from a (short) secret key by using a finite-state machine called the keystream generator, then the best one can achieve is the *practical secrecy* with respect to an adversary with limited computational resources.

For decades, the research in the stream cipher area has been classified and as such confined to the military and national security communities. However, in the last two decades, since the invention of public-key cryptography, interest for stream cipher analysis has increased in academic community and many research papers have been published. This paper is intended to provide a survey of recent results in this very interesting and sensitive area of public cryptology, with emphasis on (published and unpublished) contributions by the author. In addition, the area of secret sharing systems is discussed too. Numerous open mathematical and research problems are pointed out.

Stream Ciphers. Let $x = (x_t)_{t=0}^{\infty}$ and $y = (y_t)_{t=0}^{\infty}$ denote the plaintext and ciphertext binary sequences,[1] respectively, and let $s = (s_t)_{t=0}^{\infty}$ denote the internal state sequence where s_t is a binary vector of length M. Let the binary vectors (strings) k and r stand for the secret and randomizing keys, respectively, which determine the initial internal state $s_0(k, r)$. Then, a general binary stream cipher decipherable without delay is an invertible nonautonomous finite-state machine with one input and one output which maps an input sequence x into the output sequence y by the encryption (sequential) transform defined by

$$s_{t+1} = F_k(s_t, x_t), \quad t \geq 0 \tag{1}$$

$$y_t = x_t + f_k(s_t), \quad t \geq 0 \tag{2}$$

[1]For a sequence $a = (a_t)_{t=0}^{\infty}$, the notation $(a(t))_{t=0}^{\infty}$ is also used throughout.

where the addition is binary (modulo 2) and $F_k : \{0,1\}^{M+1} \rightarrow \{0,1\}^M$ and $f_k : \{0,1\}^M \rightarrow \{0,1\}$ are the secret key dependent next-state and output functions (see [**73**]). Typically, F_k and f_k do not depend on k. The sequence $z = (z_t = f_k(s_t))_{t=0}^{\infty}$ is called the keystream sequence. The inverse, decryption transform is defined by essentially the same equations.

The basic two modes of operation of stream ciphers are the keystream generator (KG) or the pseudorandom sequence generator mode and the self-synchronizing (S^2) or the cipher feedback mode. In the KG mode, the next-state function does not depend on the plaintext bit, that is, $s_{t+1} = F_k(s_t)$, so that the keystream sequence z is plaintext independent. To deal with synchronization errors when encrypting longer messages with the same secret key k, a long message is divided into shorter ones and a randomizing key r (typically sent in the clear) is used to reinitialize the keystream generator for every new message to be encrypted.

In the S^2 mode, the decryption transform has a finite input memory, that is, $s_{t+1} = (y_{t-i})_{i=0}^{M-1}$, so that the keystream sequence depends on ciphertext only. Its security entirely depends on the output (feedback) function f_k which must be secret key dependent.

The third mode of operation of stream ciphers is defined by general equations (1) and (2) where the next-state function effectively depends on the current plaintext symbol. In [**46**] it is called the stream cipher with memory (SCM) mode, as each ciphertext bit does not depend on the current plaintext bit only, but also on the previous plaintext bits as well. The SCM mode is typically overlooked as a practical possibility in the open literature on stream ciphers (e.g., see [**73**], [**74**], and [**65**]). The synchronization and substitution errors on real channels couuld be dealt with by separate error-correction and/or detection codes and, in addition, by the resynchronization method. However, instead of using and transmitting the randomizing key as described above, one may just prepend the randomizing key to each new message and keep the initial state secret key dependent only.

Stream ciphers are required to be practically secure with respect to computationally bounded cryptanalytic attacks in the known plaintext/ciphertext scenario. As knowing a plaintext/ciphertext string pair is equivalent to knowing a keystream/ciphertext string pair, the *practical security criterion* for a KG mode is equivalent to the *keystream unpredictibility criterion*, which means that without knowing the secret key it should be computationally infeasible to reconstruct a keystream sequence from its portions or, more generally, from portions of a set of keystream sequences obtained from different randomizing keys. Most papers in the open literature on the cryptanalysis of stream ciphers deal with the KG mode without resynchronization (see [**73**], [**17**], and [**65**] for recent surveys), several of them with the S^2 mode, and just a few with the SCM mode or the KG mode with resynchronization.

In practice, the stream cipher security is checked only with respect to particular cryptanalytic attacks, and the required immunity to these attacks gives rise to various practical design criteria. Cryptanalytic attacks can be classified into three general types. The attacks of the first type use statistical weaknesses of the keystream sequence for prediction and the resulting design criterion is known as good statistical properties of the keystream sequence. The attacks of the second type aim at reconstructing the keystream sequence by using an equivalent stream cipher of a simple structure and typically much larger internal state size whose parameters have to be defined from known portions of the keystream sequence. Such attacks have to be repeated for every new randomizing key. The corresponding derived design criteria include long period and high complexity measures of various kinds.

The attacks of the third type aim at reconstructing the secret key and as such do not have to be repeated for every new randomizing key. Most common attacks of this type have a limited goal of reconstructing the initial state or an internal state at a given time only. In the case of resynchronization, it remains to recover the secret key that controls the initial state by using the known randomizing key.

The most common type of keystream generators found in the open literature consists of a number of possibly irregularly clocked linear feedback shift registers (LFSRs) that are combined by a function with or without memory. An LFSR is a finite-state machine producing a sequence of elements from a finite field or a finite ring satisfying a linear recursion specified by the feedback polynomial. If the feedback polynomial is chosen to be primitive over a finite field, then the LFSR sequence achieves the maximum possible period for a given LFSR length and is known to possess good statistical properties measured by the distributions of certain patterns on a period (see [**54**]). However, an LFSR sequence is easily predictible. This is the reason why nonlinear combining funcitons, with or without memory, and/or irregular clocking are used.

Another known type of keystream generators have the next-state function based on slowly-varying tables instead of shift registers. Such a table consists of a number of elements from a finite set (e.g., $\{0,1\}^n$) and changes in time in such a way that only few elements are changed at each time. The best known example which is known as RC4 is used in many commercial products worldwide and is (allegedly) publicized in [**74**].

It is shown in [**46**] and [**42**] how to convert any keystream generator into the SCM mode and practical security of both the modes is discussed. *Analyzing the security of the resulting SCM mode is a new research area with many interesting problems.* In [**46**] it is proposed how to construct secure self-synchronizing stream ciphers and other cryptographic primitives such as keyed hash functions, hash functions, and block ciphers from any secure stream cipher in the SCM mode.

Period, Sequence Complexities, and Statistical Tests. Basic design criteria for keystream generators are a large period and good statistical properties of the keystream sequence. For some sequences such as the LFSR sequences, one can establish the so-called long-term statistical properties. They are measured as the frequency distribution of certain patterns on a period which should be close to the expected value for a purely random sequence (e.g., the Golomb's randomness postulates [**54**]).

The short-term statistical properties are measured by statistical tests applied to sequence segments shorter than the period. Standard statistical tests of randomness include the frequency test, the serial test, the poker test, the runs test, and the autocorrelation test (e.g., see [**3**] and [**65**]). A so-called repetition test [**55**], [**56**] enables one to consider blocks of bits that are twice as large as the blocks in the poker test. The so-called universal test [**61**] measures the entropy of a binary information source by using a universal data compression code. For Markov sources, it appears to be less powerful than the poker and the runs tests, respectively, which are also able to detect any statistical deviation in a sufficiently long sequence produced by a stationary ergodic source (see [**57**]).

Apart from stream cipher cryptography, pseudorandom sequences generated by finite-state machines have numerous applications in many other areas of computer science and electrical engineering such as spread spectrum communications, radar ranging, random number generation, computer simulations, global positioning systems, and software testing. Deriving the period of a pseudorandom sequence is generally a difficult algebraic problem which seems to be tractable only for relatively simple sequences and under special constraints. However, due to the unpredictibility criterion, keystream sequences should not have a simple structure.

Accordingly, a common design approach is to obtain the keystream sequence from some simple-structured elementary recurring sequences with controllable period which are combined by a suitable operation such as a nonlinear combining function with or without memory, nonuniform decimation, and interleaving. Given a sequence s over any set and a positive integer d, the *uniform decimation* of s by d is defined by $(s(td))_{t=0}^{\infty}$. In general, given a decimation sequence $d = (d_t)_{t=0}^{\infty}$ as a periodic nonnegative integer sequence, the (nonuniform) decimation of s by d is defined by $\left(s\left(:\ \sum_{i=0}^{t} d_i\right)\right)_{t=0}^{\infty}$. A clock-controlled LFSR can be described by nonuniform decimation. The period of interleaved and nonuniformly decimated integer sequences is analyzed in [**44**].

A high linear complexity of the keystream sequence has also become a standard design criterion especially for keystream generators based on LFSRs. It is defined as the length of the shortest LFSR that can generate a sequence. An overview of basic results on the period and the linear complexity of combined sequences obtained from the LFSR sequences can be found in [**72**], [**73**], and [**65**] (see also [**53**] for clock-controlled LFSRs). In this direction, memoryless combining functions are treated in [**9**] and a probabilistic approach for irregularly clocked LFSRs is introduced in [**8**]. A specific scheme based on LFSRs and a slowly-varying table is analyzed in [**10**]. The number of output sequences generated by this scheme from all possible initial states is derived in [**12**]. The periods of the so-called multiplexed sequences and their generalizations are determined in [**36**]. A novel approach for finding the minimum polynomial of uniformly decimated sequences over an arbitrary field is given in [**25**].

Apart from the linear complexity, one can analogously define the nonlinear complexity of a specified or unspecified algebraic order in terms of nonlinear recursion functions (i.e., shift registers with nonlinear feedback). Another generalization is to consider a sequence over a field F and the linear complexity over a larger field F', $F \subset F'$ (see [**58**]). One can also define complexity measures with respect to nonlinear recursion functions with memory such as the 2-adic or the p-adic complexity defined in terms of the so-called feedback with carry shift register (FCSR) [**59**].

Correlation Analysis of Regularly Clocked Combiners. To achieve the keystream unpredictibility (in particular, to increase the linear complexity), regularly clocked LFSRs are commonly combined by memoryless functions or by functions with memory. The resulting keystream generator is called a combiner with or without memory. A general binary combiner with M bits of memory and N inputs is a nonautonomous finite-state machine (sequential circuit) defined by $s_{t+1} = F(s_t, X_t)$ and $z_t = f(s_t, X_t)$, $t \geq 0$, where $F : \{0,1\}^N \times \{0,1\}^M \to \{0,1\}^M$ is the next-state vector boolean function, $f : \{0,1\}^N \times \{0,1\}^M \to \{0,1\}$ is the output boolean function, s_t is an M-dimensional internal state vector, X_t is an N-dimensional binary input vector, and z_t is the output bit (at time t). In the correlation analysis a probabilistic model in which the inputs are assumed to be mutually independent purely random binary sequences is considered. The objective is to investigate the statistical dependence between the output and input sequences. Such a dependence may be a basis for a divide-and-conquer cryptanalytic attack on the initial states of a subset of the LFSRs which is called a correlation attack.

It is shown in [**79**] and [**78**] that a memoryless combiner ($M = 0$) is vulnerable to a correlation attack based on the termwise correlation between the keystream sequence and a subset of the LFSR sequences. Defining the correlation coefficient between two binary random variables x and y as $c(x, y) = 2\Pr\{x = y\} - 1$, it is pointed out in [**63**] that for every boolean function f, the sum of the squares of the correlation coefficients between the output bit and all linear functions of the input bits is equal to 1. Any nonzero linear (or nonlinear) function correlated to f can be used to mount a correlation attack on the involved LFSRs. The divide-and-conquer effect can be reduced by using a correlation immune boolean function [**78**], whose output is statistically independent of any specified number (called the correlation immunity order) of inputs. There is a trade-off between the linear complexity and the correlation immunity [**78**]. To overcome this trade-off, it is suggested in [**72**] to use combiners with memory. It is shown that the maximum-order correlation immunity (with respect to input sequences) can be achieved, regardless of the linear complexity, by using only one bit of memory. To this end, the so-called summation generator is proposed where the LFSR sequences are regarded as binary expansions of integers (more precisely, as 2-adic numbers) and the output sequence is formed by integer summation (with carry). Correlation properties of combiners with one bit of memory are further investigated in [**64**].

Correlation properties of a general binary combiner with an arbitrary number M of memory bits are derived in [**30**]. For any positive integer m, the sum of the squares of the correlation coefficients between all nonzero linear functions of m consecutive output bits and all linear functions of the corresponding m consecutive inputs is shown to be dependent upon a particular combiner, unlike the memoryless combiners. The minimum and maximum values of the correlation sum as well as the necessary and sufficient conditions for them to be achieved are determined. As every linear function when applied to a set of sequences defines a linear sequential transform, with a finite input memory, of the set, it follows that in combiners with memory there exists the termwise correlation between linear transforms of the output sequence and linear transforms of the input sequences. *Interesting open problems are to find constructions for nonlinear combiners with memory with the minimum correlation sum as well as to to examine the conditions under which uniform or approximately uniform distribution of the correlation among all pairs of output and input linear functions is achievable.*

In [**16**] and [**30**] an efficient linear sequential circuit approximation method for finding output and input linear transforms with comparatively large correlation coefficients which is feasible for large M is developed. The method consists in deriving and solving a linear sequential circuit with additional nonbalanced (nonuniformly distributed) inputs that is based on linear approximations of the boolean output and component next-state functions. The corresponding correlation attack on combiners with LFSRs is proposed and it is shown that every such combiner is essentially zero-order correlation immune. It turns out that the security of combiners with memory can be considerably improved by increasing M.

Keystream Statistics and Linear Models. Many proposed keystream generator schemes seem to possess good statistical properties with respect to the standard statistical tests. If the keystream sequence is produced from a set of elementary sequences with good statistical properties (such as maximum-length sequences), then the keystream statistics can also be analyzed theoretically by an approach proposed in [**30**]. Namely, assume that the elementary sequences are purely random and mutually independent (if applicable) and analyze if the keystream sequence is also purely random. One thus verifies if the combining function (possibly including irregular clocking) spoils the good statistics of the elementary sequences. *An open problem is to establish necessary and sufficient conditions for the output sequence of a general binary combiner with memory to be purely random and thus generalize the sufficient conditions given in* [**30**]. The most illustrative examples of how a combining function may spoil the statistics are the well-known multiplexer generator and the nonlinear filter generator. The multiplexer generator consists of two binary LFSRs where k chosen stages of one LFSR determine which out of 2^k chosen stages of the second LFSR is selected to give the output bit at any given time (see [**3**]). It is used for video encryption in the European standard for pay-television. A detectable autocorrelation weakness of the multiplexer generator is established in [**18**] and [**49**]. It can be regarded as a linear statistical weakness described below.

The nonlinear filter generator is the simplest shift register based keystream generator. Its output sequence is obtained by a nonlinear sequential transform of only one LFSR sequence, that is, by applying a memoryless function to different phase shifts of the same LFSR sequence. Let $x = (x(t))_{t=-r}^{\infty}$ be a binary maximum-length sequence of period $2^r - 1$ ($(x(t))_{t=-r}^{-1}$ is the LFSR initial state), let $f(y_1, y_2, \ldots, y_n)$ be a boolean function of n, $n \leq r$, nondegenerate input variables, and let the tapping sequence $\gamma = (\gamma_i)_{i=1}^n$ be an increasing sequence of nonnegative integers such that $\gamma_1 = 0$ and $\gamma_n \leq r - 1$. Then the output sequence $z = (z(t))_{t=0}^{\infty}$ of the nonlinear filter generator is defined by $z(t) = f(x(t - \gamma_1), \ldots, x(t - \gamma_n))$, $t \geq 0$. It can be viewed as a combiner with one input sequence and with a finite input memory of $M = \gamma_n - \gamma_1$ bits. Assuming that the input sequence is purely random, it is shown in [**28**] that the output sequence is purely random for every γ if the filter function is linear in the first or the last variable. *A related open problem is to find the corresponding necessary and sufficient conditions.*

Statistical properties of the keystream sequence should also be investigated by using the statistical tests adapted to the structure considered. It is proposed in [**20**] and [**29**] to study linear statistical weaknesses of the binary keystream sequence defined as linear equations satisfied by the keystream bits with probability, p, different from one half, that is, with the correlation coefficient, c, different from zero, where $c = 2p - 1$ (assuming that the initial state is chosen uniformly at random). If a binary keystream generator has the internal memory size of M bits, then such a weakness necessarily exists on any block of $M + 1$ consecutive keystream bits [**20**] and can be detected on a keystream segment of length $O(1/c^2)$ (for small c). Accordingly, it is an effective statistical weakness if c is (much) bigger than $2^{-M/2}$. As a consequence, if the next-state function is one-to-one, then the generator can be linearly modeled as a nonautonomous LFSR of length at most M with an additive input sequence of nonbalanced identically distributed binary random variables. The sum of the squares of the correlation coefficients over all the linear models of any given length proves to be dependent on a keystream generator. The minimum and maximum values of the correlation sum along with the necessary and sufficient conditions for them to be achieved are established in [**29**]. *Constructions of nonlinear keystream generators with the minimum correlation sum as well as with uniform or approximately uniform distribution of the correlation among all linear models of a given length remain to be investigated.*

An effective method for the linear model determination based on the linear sequential circuit approximation of autonomous finite-state machines is proposed in [**29**] too. It is thus shown how to construct linear models with comparatively large correlation coefficients for clock-controlled LFSRs, for combiners with or without memory, and for arbitrary keystream generators based on regularly or irregularly clocked shift registers with linear or nonlinear feedback. Keystream generators based on slowly-varying tables may also suffer from linear statistical weaknesses. For example, a linear model for the well-known RC4 keystream generator is established in [**39**] and [**48**].

Any linear model is the basis for a structure-dependent and initial-state-independent statistical test. As such, it can be used either directly for reconstructing a statistically redundant plaintext or for reconstructing the secret key controlling the structure of the generator. For example, it is proposed in [**23**] how to reconstruct the unknown feedback polynomial of a clock-controlled LFSR. Multiple linear models may be used for the initial state reconstruction. In addition, replacing parts of a keystream generator by their linear models may result in the correlation attacks on the initial state of the remaining parts. As the linear sequential circuit approximation of autonomous finite-state machines yields linear models as well as mutually correlated linear transforms of the output sequence and of the internal sequences depending on parts of the internal state only (to be used in correlation attacks), it is called in [**21**] the linear cryptanalysis of stream ciphers. A keystream generator that should be immune to linear cryptanalysis is proposed in [**21**].

Fast Correlation Attacks on Regularly Clocked Linear Feedback Shift Registers. Consider a combiner with or without memory where a linear transform of the input LFSR sequences (which itself is also an LFSR sequence) is found to be termwise correlated to a linear transform of the output sequence (called here the observed keystream sequence) with the correlation coefficient $c > 0$. Then the observed keystream sequence of length N, z^N, can be modeled as the output sequence of a memoryless binary symmetric channel (BSC) with error probability $p < 0.5$ ($c = 1 - 2p$) when the unknown LFSR sequence of length N, a^N, is applied to its input. The LFSR feedback polynomial $f(x)$ of degree r is assumed to be known. The set of sequences a^N generated from all the LFSR initial states is then a linear (N, r) code. The basic correlation attack [**79**], with computational complexity $O(2^r)$, is then the minimum Hamming distance (optimal) decoding. The decoding error probability will be close to zero if $r/N < C$ where $C = 1 - H_2(p)$ is the capacity of the BSC (for small c, if $N > r\,O(1/c^2)$).

The objective of fast correlation attacks, first introduced in [**62**], is to recover the original LFSR sequence without searching over all 2^r LFSR initial states. This can be achieved by using iterative probabilistic decoding procedures aiming at minimizing the bit-error rate. They are based on the parity-checks defined from the phase shifts of the parity-check polynomials obtained as the polynomial multiples $h(x)$, $h(0) = 1$, of $f(x)$ of low weight (number of nonzero terms) and of as small degree as possible. If the weight of $f(x)$ is low, then the low-weight parity-check polynomials can be obtained by repeated squaring of $f(x)$ [**62**], and if not, then one can use the polynomial residue method [**34**].

For a random $f(x)$, it is argued in [**34**] that the expected minimum degree of the polynomial multiple $h(x)$ of any given weight w, $w \geq 2$, is $O(2^{r/(w-1)})$. *An interesting open problem is to examine the case when $f(x)$ is randomly chosen with a specified weight.*

The iterative probabilistic decoding algorithm [**69**], [**68**], [**83**] with a modification given in [**26**] consists of several rounds, each composed of a number of iterations. For each of N observed keystream bits, a set of (preferably orthogonal) parity-checks is first determined. The algorithm starts with the observed keystream sequence z^N and with p as the error probability for each bit. The sequence z^N is then iteratively modified to yield the reconstructed LFSR sequence. In each iteration the parity-check values are recalculated and the current error probabilities are computed as the posterior probabilities of error given the previous error probabilities as the prior probabilities of error. Then all the bits with the error probability bigger than 0.5 are complemented. If p is not too close to 0.5, then most error probabilities quickly converge to zero. By using a probabilistic argument, a practical convergence condition yielding the required keystream sequence length is derived in [**70**] (see [**83**] for other conditions). *It remains to study the convergence problem by analyzing the fixed points of the underlying nonlinear operator, which seems to be difficult.*

In order to correct all the errors, the algorithm is repeated for several rounds each time resetting all the error probabilities to p. At the end a simple information set decoding technique which consists in searching for an error-free sliding window of r consecutive bits is used. Somewhat better performance is achieved by an improved algorithm [**26**] with the so-called fast resetting and with the sliding window technique incorporated in rounds. Other modifications including partial resetting and simulated annealing are also suggested in [**26**]. A comparison with the so-called 'free energy minimization' algorithm based on the hidden Markov chain decoding is given in [**6**].

One may also use a more sophisticated information set decoding technique where information sets are randomly chosen from a set of bits whose error probabilities are most different from p after a few iterations. The fast correlation attack may be reduced to the information set decoding stage only if a combiner with memory allows one to find sufficiently many bit positions in the LFSR sequence with the error probability significantly different from one half when conditioned on the observed keystream sequence (see [**64**] for the summation generator with two inputs). Such an attack is called the conditional correlation attack. The connection between this attack and random linear codes is established in [**35**].

In a combiner with memory, there may exist different linear transforms of the same subset of input LFSR sequences that are correlated to the same linear transform of the output sequence. In this case, if successful, the iterative algorithm converges (randomly) to an LFSR sequence corresponding to one of these input linear transforms. This phenomenon is studied in more detail in [**45**] for the summation generator and in [**41**] for the nonlinear filter generator. *A solution to the problem of finding mutually correlated input and output linear transforms in the summation generator with an arbitrary number of input LFSRs is presented in* [**45**], *but a conjecture made in this regard remains to be investigated.* It is proved in [**28**] that the practical security against the conditional correlation attack on the nonlinear filter generator can be achieved by using a full positive difference set for the tapping sequence and a correlation immune nonlinear filter function.

A fast and efficient procedure for finding approximations of low algebraic order to boolean functions depending on a large number of variables, if such approximations exist, is developed in [**31**]. The procedure uses iterative error-correction algorithms for fast correlation attacks and is based on representing low order boolean functions by appropriate linear recurring sequences generated by binary filter generators. *If the algebraic order is bigger than* 1, *then it is an open problem to determine* *low-weight binary polynomials whose roots have the form* α^e*, where* α *is a primitive element in a finite field of characteristic* 2 *and the number of* 1*'s in the binary representation of a positive integer* e *equals the given algebraic order.*

A stop/go LFSR is a clock-controlled LFSR that is at each time, according to the value of the clock-control bit, either clocked once or not clocked at all (in which case the last bit produced is repeated). An up/down LFSR is defined similarly with a difference that it is clocked one step either forwards or backwards at each time. If the clock-control sequence is bitwise added to the output sequence, then the repetition weakness existing in both stop/go and up/down LFSRs is removed. A cascade connection of such stop/go (up/down) stages where the input to the first stage is a regularly clocked LFSR sequence is called a stop/go (up/down) cascade (e.g., see [**53**]). It is proved in [**67**] that the first (second) binary derivatives of the input LFSR sequence and the output sequence of a stop/go (up/down) cascade consisting of k stages are unconditionally bitwise correlated with the correlation coefficient $1/2^k$. This can be used to mount a fast correlation attack on the input (regularly clocked) LFSR, and then successively on the remaining LFSRs in the cascade too. More importantly, a specific conditional correlation attack on one stop/go stage is proposed in [**66**] and then extended to stop/go cascades in [**60**]. A similar conditional correlation attack on up/down cascades is proposed in [**67**]. *In the theoretical analysis of the underlying Markov chains there are several open problems related to the success of conditional correlation attacks.*

Embedding and Probabilistic Correlation Attacks on Irregularly Clocked Shift Registers. Irregular clocking of LFSRs is another interesting operation aiming at achieving the keystream unpredictibility (in particular, high linear complexity, long period, and immunity to fast correlation attacks). If $x = (x(t))_{t=0}^{\infty}$ is the output sequence of a regularly clocked LFSR and if $d = (d_t)_{t=0}^{\infty}$ is a decimation (nonnegative integer) sequence produced by a finite-state machine called a clock-control generator, then the output sequence of the clock-controlled LFSR is defined as the decimated sequence $z = (z(t))_{t=0}^{\infty} = \left(x\left(\sum_{i=0}^{t} d_i\right)\right)_{t=0}^{\infty}$. Thus, in order to obtain the current output symbol $z(t)$ one has to delete $d_t - 1$ consecutive symbols from x if $d_t \geq 1$ or has to repeat $z(t-1)$ if $d_t = 0$. If $\mathcal{D}$ denotes the range of d, then the LFSR is said to be $\mathcal{D}$-clocked. In particular, if $\mathcal{D} = \{0, 1\}$, then the LFSR is stop/go clocked, if $\mathcal{D} = [1, d+1] = \{1, 2, \ldots, d+1\}$, then the irregular clocking is called constrained, and if $\mathcal{D} = [1, \infty)$, then the irregular clocking is called unconstrained.

It is assumed that the secret key controls the LFSR and the clock-control generator initial states. The clock-control generator initial state can be recovered by the exhaustive search, regardless of the LFSR initial state, by using the linear consistency test [**82**], as only the correct guess about the decimation sequence gives rise to consistent linear equations when known output bits are expressed in terms of the LFSR initial state. The objective of correlation attacks is to reconstruct the LFSR initial state without knowing the decimation sequence, from a given segment of the output sequence. There are two types of correlation attacks: embedding and probabilistic attacks. In both of them, for each assumed LFSR initial state, a segment of length $m(n)$, $x^{m(n)}$, of the regularly clocked LFSR sequence is produced and then tested by using a segment of length n, z^n, of the known output sequence. If the decimation sequence can be modeled as a random sequence, then $m(n)$ is chosen so that the (missing event) probability that the given length n is obtained from a length $m > m(n)$ is small. For constrained clocking, one can choose $m(n) = n(d+1)$. For unconstrained clocking with independent deletions, the decimation sequence is modeled as a sequence of independent identically distributed nonnegative integer random variables with a geometric probability distribution $\big(p^{i-1}(1-p)\big)_{i=1}^{\infty}$ (i.e., each bit from a regularly clocked LFSR sequence is deleted with (deletion) probability p independently of other bits).

In the embedding attack one checks if it is possible to obtain z^n as the decimation of a prefix of $x^{m(n)}$, i.e., if z^n can be $\mathcal{D}$-embedded into $x^{m(n)}$, and accepts all the LFSR initial states allowing the embedding. In the probabilistic attack one computes the (edit) probability that z^n is produced as the decimation of a prefix of $x^{m(n)}$ according to a given probability distribution of the decimation sequence and accepts all the LFSR initial states with sufficiently large probability. The attacks are successful if the number of candidate LFSR initial states obtained is small. The correct LFSR initial state along with the clock-control generator initial state are then found typically faster than by the exhaustive search.

If one defines the $\mathcal{D}$-constrained edit distance between z^n and x^m as the minimum number of deletions and effective substitutions needed to obtain z^n from x^m, where the deletions are subject to $\mathcal{D}$ except at the end, then the embedding is possible if and only if the edit distance is equal to $m-n$ (no effective substitutions). For unconstrained clocking, the edit distance coincides with the Levenshtein distance without insertions. For both constrained and unconstrained clocking, the edit distance can be computed by recursive algorithms with computational complexity $O(n(m-n))$ (see [**13**] for constrained clocking). If the decimation sequence is assumed to be a sequence of independent identically distributed random variables, then the edit probabilities can also be computed by similar recursive algorithms (e.g., the noiseless version of the constrained clocking algorithm in [**15**] and the unconstrained clocking algorithm in [**22**]).

Define a $\mathcal{D}$-embedding probability $P_{\mathcal{D}}(z^n, m)$ as the probability that a given binary string z^n of length n can be $\mathcal{D}$-embedded into a purely random (uniformly distributed) binary string x^m of length m. The embedding attack can be made successful for any LFSR by choosing a sufficiently large n if and only if $P_{\mathcal{D}}(z^n, m(n))$ tends to zero when n increases, for all z^n or for almost all z^n (i.e., for a fraction of all z^n tending to 1 as n increases). If the embedding probability decreases exponentially, then the minimum necessary keystream length, n, for the successful attack is linear in the shift register length.

In [**84**] a constrained embedding attack on a binary $[1,2]$-clocked shift register is introduced and an exponentially small upper bound on $P_{[1,2]}(z^n, 2n)$ holding for every z^n is established. The upper bound is improved in [**33**] by solving the underlying combinatorial problem. For $d > 1$, exponentially small upper bounds on

$P_{[1,d+1]}(z^n, n(d+1))$ holding for almost all z^n are derived in [**22**] and [**27**], but the bounds are not tight. *Deriving a tighter upper bound on the constrained embedding probability holding for all (or almost all)* z^n *remains an open problem for* $d > 1$. For unconstrained clocking with independent deletions with probability p, it is proved in [**22**] that the unconstrained embedding attack is successful if and only if $p < 0.5$.

Probabilistic correlation attacks for constrained clocking and unconstrained clocking with independent deletions are proposed in [**15**] and [**22**], respectively. Such an attack is successful if and only if the capacity of the corresponding communication channel with deletion errors is positive, which is conjectured in [**22**]. *Verifying this conjecture as well as deriving the capacity of such channels seem to be difficult problems.* For unconstrained clocking with independent deletions with $p = 0.5$, successful experimental probabilistic attacks are conducted and reported in [**80**].

Edit Distance and Edit Probability Correlation Attacks on Irregularly Clocked Combiners. Several clock-controlled LFSRs can be combined by a linear or nonlinear function with or without memory into a more complex scheme. One can then generally define the (divide-and-conquer) correlation attacks whose objective is to reconstruct the initial states of a subset of the LFSRs without knowing the individual decimation sequences and the initial states of the remaining LFSRs. The edit distance correlation attack is based on an appropriate edit distance which is defined in terms of the edit transformation consisting of deletions of symbols from regularly clocked LFSR sequences and of substitutions of symbols in the combination sequence obtained from the resulting decimated sequences by a function with or without memory, which is derived from the combining function of the combiner considered. The edit distance is defined as the minimum number of deletions and effective substitutions needed to obtain a given output string from assumed input strings by this edit transformation. If the optimum edit transformation for correctly guessed LFSR initial states does not include effective substitutions (the noiseless case), then the edit distance correlation attack can be called the (generalized) embedding attack. The edit distance correlation attack on a single shift register for the noisy case with constrained clocking is first introduced in [**11**] and [**13**] along with a recursive algorithm for the edit distance computation. The attack applies to a binary memoryless combiner with a zero-order correlation immune combining function in which the output sequence is bitwise correlated to a clock-controlled LFSR sequence (to be reconstructed). More generally, recursive algorithms for computing the edit distances for constrained clocking are derived in [**14**] for memoryless combining functions and in [**32**] for combining functions with memory.

The edit probability (probabilistic) correlation attack is based on an appropriate edit probability which is defined in terms of the (random) edit transformation consisting of deletions of symbols from regularly clocked LFSR sequences according to a given probability distribution of the decimation sequences, of combining the resulting decimated sequences by a function with or without memory, and of substitutions of symbols in the combination sequence only in the noisy case when these substitutions are required for correctly guessed LFSR initial states. The substitutions appear if the objective is to recover the initial states of a subset of the LFSRs and when the effect of the remaining LFSRs can be modeled by a correlation noise. The edit probability is defined as the probability that assumed input strings yield a given output string by this edit transformation. The edit probability correlation attack on a single shift register, for the same scheme as in [**13**], is first introduced in [**15**] along with a recursive algorithm for the edit probability computation. Effective substitutions are assumed to take place independently with the same probability corresponding to the correlation coefficient used. The computation of edit probabilities usually takes more time and/or space than the computation of edit distances, but the resulting correlation attack is more powerful (typically, statistically optimal).

To increase the divide-and-conquer effect of a correlation attack, by using a linear transform of the output sequence that is correlated to a linear transform of a subset of the input sequences, we obtain that a linearly transformed output sequence of the given combiner with memory is correlated to the output sequence of a linear combiner with a finite input memory which is applied to a subset of the clock-controlled LFSR sequences to be reconstructed in the correlation attack. The corresponding edit distances and probabilities as well as the recursive algorithms for their efficient computation for both constrained and unconstrained irregular clocking are given in [**37**] and [**51**], respectively.

The success of generalized embedding correlation attacks can be analyzed by the corresponding generalized embedding probabilities essentially in the same way as for a single clock-controlled shift register. *Determining these probabilities for combining functions with or without memory are interesting combinatorial problems.* On the other hand, the success of edit distance correlation attacks in the noisy case and edit probability correlation attacks is based on the expectation that the edit distance and the edit probability for correctly guessed LFSR initial states are close to being minimal and maximal, respectively. In the underlying statistical hypothesis problems one should consider the probability distributions of the edit distance and the edit probability in the two cases: when the LFSR initial states are guessed correctly and incorrectly, where the latter one can further be modeled by the input strings being statistically independent of the given output string. *Theoretical analysis of the statistical discrimination between the two distributions appears to be a difficult problem. Another approach to analyzing the success of the edit probability correlation attack is through the capacity of the corresponding communication channel with deletion and substitution errors, but this seems to be difficult too.*

With respect to stop/go irregular clocking, correlation attacks on a keystream generator known as the alternating step generator (ASG) that are based on specific edit distances and edit probabilities are proposed in [**40**] and [**50**], respectively. The ASG output is obtained as the binary sum of two binary LFSRs, LFSR_1 and LFSR_2, that are stop/go and go/stop clocked, respectively, according to the third, regularly clocked LFSR. The edit distance correlation attack [**40**] targets the initial states of LFSR_1 and LFSR_2 combined, whereas the edit probability correlation attack [**50**] targets the initial states of individual LFSRs. The corresponding embedding probability for the edit distance attack is experimentally shown to be exponentially small in the output string length. *Whether the embedding probability can be determined theoretically remains to be investigated.* A similar result is obtained for the false alarm probability corresponding to the edit probability attack. *Establishing this theoretically seems to be a difficult problem.*

If the output sequence of an irregularly clocked combiner with memory is irregularly interleaved with another pseudorandom sequence to form the keystream sequence, it is then possible to introduce the correlation attacks based on edit distances or edit probabilities corresponding to edit transformations that incorporate the insertion of symbols in the combination sequence. Such edit distances along with the recursive algorithms for their computation are introduced in [**71**] for a single clock-controlled shift register and in [**24**] for an irregularly clocked combiner.

The correlation attacks described above require the exhaustive search over the initial states of the involved LFSRs and are hence feasible only if the effective secret key controlling these initial states is short. *One of the most interesting problems in the cryptanalysis of stream ciphers is to develop faster correlation attacks on irregularly clocked shift registers. One way of doing this may be by adapting or modifying the edit probability correlation attacks. Another way may be based on the method for a single clock-controlled LFSR whose theoretical framework is proposed in* [**23**]. *It essentially consists in an iterative probabilistic deletion-error-correction algorithm that is based on specific low-weight parity-checks and remains to be experimentally studied.*

Internal State Reversion Attacks and Branching Processes. Two internal state reversion attacks are presented in this section. The objective of one of them which is applied to a keystream generator known as (alleged) A5 is to recover the initial state from a known internal state at a given time. The next-state and the output functions are assumed to be known. As a cryptanalytic method it may also be applicable to any keystream generator whose next-state function is not one-to-one (the reversion is trivial for one-to-one next-state functions). The other attack is applied to the nonlinear filter generator and has the objective of recovering the whole internal state at a given time from an assumed part of the internal state. Alternatively, it can also be viewed as an inversion attack on a combiner with one input, one output, and with a finite input memory whose objective is to recover a segment of the input sequence from a known segment of the output sequence. Both attacks are described in terms of the theory of critical and/or subcritical branching processes [**1**].

The keystream generator known as A5 is used for stream cipher encryption in the GSM standard for digital cellular mobile telephones. According to [**74**], A5 consists of three binary LFSRs with known primitive feedback polynomials that are mutually clocked in the stop/go manner. Middle taps in each of the LFSRs define the clock-control bits the majority of which determine which LFSRs are clocked at each time. The keystream sequence is formed as the bitwise sum of the three stop/go clocked LFSR sequences. The secret (session) key is nonlinearly combined with a known randomizing key (frame number) to form the LFSR initial states. More precisely, the LFSR states are first defined by the secret key and the randomizing key is then bitwise added into the feedback path of each of the LFSRs that are mutually clocked as above. The last LFSR states represent the initial LFSR states for the keystream generation. A number of the firstly produced output bits are discarded and the keystream sequence generated is very short.

The cryptanalytic approach proposed in [**38**] and [**52**] consists of several methods for internal state reconstruction, initial state reconstruction, and secret key reconstruction, where the internal state consists of the three LFSR internal states. The internal state reconstruction can be carried out either by generating and solving a specific set of linear equations or by a method based on a time-memory trade-off. The initial state reconstruction and the secret key reconstruction can be achieved by the internal state reversion based on the theory of branching processes. *An interesting theoretical problem is to develop edit distance or edit probability correlation attacks on A5 without the restriction on the keystream sequence length available.*

The internal state reversion attack recovers all the initial states that result in a given internal state at a given time and consists of the recursive computation and storage of the reverse next-state function, which is shown to be of the many-to-one type. The internal states obtained are stored as nodes in a tree. Depending on the time considered, the output sequence is assumed to be known or unknown. The secret key candidates are obtained from the initial state candidates by a similar internal state reversion attack assuming that the output sequence is unknown and that the randomizing key is incorporated into the next-state function. The obtained candidates are then checked on a small number of additional keystream sequences to find the correct secret key.

The time and space complexities of the internal state reversion attacks are determined by the size of the resulting trees. It is proposed in [**38**] and [**52**] that the trees can be analyzed by the theory of branching processes. By deriving the corresponding branching probability distributions, it is shown that the associated Galton-Watson branching process (see [**1**]) is critical in the cases when the output sequence is not known and when the randomizing key is incorporated and subcritical in the case when the output sequence is known.

Consider now the binary nonlinear filter generator described in Section 8. The objective of the inversion attack [**28**], [**47**] is to recover the unknown input sequence from a known segment of the output sequence, provided that the filter function f, the tapping sequence γ, and the LFSR feedback polynomial are known.

The inversion attack [**28**] applies to the case when f is linear in the first or the last input variable and runs forwards or backwards accordingly. Its time complexity is $O(2^M)$, where M is the input memory size. The generalized inversion attack [**47**] applies to the general case of an arbitrary f.

It is shown in [**47**] that the theory of critical branching processes can be applied to analyze the time and space complexities of the generalized inversion attack. Both theory and systematic experimental results obtained show that, almost regardless of the LFSR length, f, and γ, the time complexity is close to $O(2^M)$, with a relatively small additional space required to store the corresponding binary trees.

The attack can be extended to deal with more than one bit at a time in which case the resulting trees are no longer binary, but the associated branching process remains critical. *Construction of the filter functions maximizing the time and/or space complexities of the inversion attack in this more general setting seems to be an interesting research problem.*

Secret Sharing Schemes and Matroids Secret sharing schemes are independently proposed in [**74**] and [**4**] in order to accomplish shared control of critical actions and safeguarding cryptographic keys. A secret sharing scheme is a procedure of sharing a secret key k from a finite set $\mathcal{K}$ by distributing shares from a finite set $\mathcal{S}$ among a finite set $\mathcal{P}$ of participants.

Let the extended set of participants be defined as $\mathcal{P}^* = \mathcal{P} \cup D$, where D is a dealer, let $N = |\mathcal{P}^*|$, and let $\mathcal{R}$ denote a finite randomizing set. Further, for any function $F : \mathcal{R} \times \mathcal{P}^* \to \mathcal{K} \cup \mathcal{S}$ and arbitrary subsets $R \subseteq \mathcal{R}$ and $P \subseteq \mathcal{P}^*$, let for any $r \in \mathcal{R}$, $F(r, P)$ denote the $|P|$-tuple $(F(r,p))_{p \in P}$ and let $F(R, P) = \{F(r, P) : r \in R\}$. Then a secret sharing scheme is defined in terms of a function (matrix indexed by r and p) F such that $F(\mathcal{R}, D) = \mathcal{K}$ and $F(\mathcal{R}, \mathcal{P}) \subseteq \mathcal{S}^{|\mathcal{P}|}$.

When the dealer D wants to distribute shares corresponding to a secret key, D first randomly picks a key k according to an arbitrary prior probability distribution on $\mathcal{K}$, uniformly at random chooses a value r such that $F(r, D) = k$, and then distributes the share $F(r, p)$ to a participant p, for every $p \in \mathcal{P}$. A secret sharing scheme $\mathcal{F}$ is then uniquely determined by F and is in fact an N-tuple of discrete random variables. If no two rows of F are identical, then a scheme is called canonic.

Let Γ be a monotone set of subsets of $\mathcal{P}$. A secret sharing scheme $\mathcal{F}$ is called perfect with an access structure Γ if for every prior probability distribution of secret keys: (1) for any qualified subset of participants $P \in \Gamma$, $H(D|P) = 0$ and (2) for any unqualified subset of participants $P \notin \Gamma, P \subseteq \mathcal{P}$, $H(D|P) = H(D)$, where H is the entropy operator. The information rate for a secret sharing scheme is defined as $\log_2 |\mathcal{K}| / \log_2 |\mathcal{S}|$. A perfect secret sharing scheme is said to be ideal if it has the maximum information rate 1. Let then $\mathcal{S} = \mathcal{K}$ and let $|\mathcal{K}| = q$.

A general problem related to secret sharing schemes is to find a secret sharing scheme with minimum information rate for a given monotone access structure or, alternatively, to characterize all achievable monotone access structures for secret sharing schemes with a given i nformation rate. It is known how to construct a perfect secret sharing scheme for an arbitrary monotone access structure. Moreover, it is proved that a certain access structure cannot be realized by ideal schemes. A connection between ideal secret sharing schemes and matroids is established in [**5**]. It is proved that subsets of participants with independent shares (in the information-theoretic sense) constitute independent sets of an associated matroid. *Characterizations of the associated matroids and the achievable access structures as* *well as a general method for the construction of ideal secret sharing schemes for an arbitrary q are interesting related problems.* It is proved in [**75**] that the well-known Vamos matroid (for $N = 8$) cannot be associated with any ideal scheme.

In [**43**] a characterization of ideal schemes for an arbitrary key size q is derived in terms of balanced maximum-order correlation immune functions. A surjective function $f : \mathcal{A} \to \mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are finite sets, is called balanced if every value from $\mathcal{B}$ is assumed an equal number of times by f. A balanced function $f : \mathcal{K}^m \to \mathcal{K}$ is called maximum-order correlation immune [**78**] if it is balanced for each fixed value of every proper subset of its m input variables. For $q = 2$, it is proved in [**43**] that a matroid is an associated matroid for a binary ideal scheme if and only if it is representable over the binary field. It is independently proved in [**2**] that a similar result holds for $q = 3$ too. *Characterizing the associated matroids for $q \geq 4$ remains an open problem.* Computationally efficient access structure characterization of binary ideal schemes is established and a general method for their construction is also pointed out. *An open problem is whether this characterization can be further simplified.*

The rank function of a polymatroid on a finite set E is a real-valued set function ρ defined on subsets of E that is nondecreasing ($\rho(X) \leq \rho(Y)$, $X \subseteq Y$), submodular ($\rho(X \cup Y) + \rho(X \cap Y) \leq \rho(X) + \rho(Y)$), and satisfies $\rho(\emptyset) = 0$ (e.g., see [**81**]). An integer-valued ρ such that $\rho(X) \leq |X|$ for every subset X is the rank function of a matroid. Let $\mathcal{F}$ be a canonic secret sharing scheme on an extended set of participants $\mathcal{P}^*$. Let $H(P)$, $P \subseteq \mathcal{P}^*$, denote the joint entropy set function associated with $\mathcal{F}$. It is known that H is a rank function of a polymatroid on $\mathcal{P}^*$. Interestingly, it is not yet known if H can be equal to the rank function of any polymatroid. We conjecture that this is true asymptotically, for $|\mathcal{K}|$, $|\mathcal{S}|$, and $|\mathcal{R}|$ arbitrarily large. This is proved in [**19**] for $N = 2$ and $N = 3$.

CONJECTURE: *Let ρ be the rank function of an arbitrary polymatroid on a set of participants $\mathcal{P}^*$ such that $\rho(\mathcal{P}^*) = 1$. Then for every $\varepsilon > 0$, there exists a canonic secret sharing scheme on $\mathcal{P}^*$ such that $|H(P)/H(\mathcal{P}^*) - \rho(P)| < \varepsilon$ for every $P \subseteq \mathcal{P}^*$.*

Conclusions. The most important design strategy for a stream cipher should be the resistance to known and foreseeable cryptanalytic attacks of various types. Although an efficient attack is typically adapted to a concrete structure considered, most attacks are based on more general cryptanalytic methods and principles which apply to a wider class of stream ciphers. For some attacks, the resistance can be achieved by satisfying a set of derived design criteria. More generally, an attack is ineffective if it cannot be successfully applied experimentally.

Developing various cryptanalytic principles, methods, and concrete attacks is necessary for achieving a reasonable level of scientific rather than just intuitive confidence in the practical security of stream ciphers. This requires a creative application of various branches of mathematics, computer science, and electrical engineering including algebra, probability and statistics, coding and information theory, combinatorics, discrete optimization, and theory of algorithms. The main objective of this paper is to provide an overview of recent results and open problems and thus promote research in this relatively new scientific field.

References

[1] K. B. Athreya and P. E. Ney, *Branching Processes*, Springer-Verlag, Berlin, 1972.

[2] A. Beimel and B. Chor, *Universally ideal secret sharing schemes*, IEEE Trans. Inform. Theory **40** (1994), 786–794.

[3] H. Beker and F. Piper, *Cipher Systems: The Protection of Communications*, Wiley, New York, 1982.

[4] G. R. Blakley, *Safeguarding cryptographic keys*, *Proc. AFIPS 1979 NCC*, vol. 48, New York, 1979, 313–317.

[5] E. F. Brickell and D. M. Davenport, *On the classification of ideal secret sharing schemes*, J. Cryptology **4** (1991), 123–134.

[6] A. Clark, J. Dj. Golić, and E. Dawson, *A comparison of fast correlation attacks*, *Fast Software Encryption – Cambridge '96* (D. Gollmann, ed., Springer-Verlag), Lecture Notes in Computer Science **1039** (1996), 145–157.

[7] W. Diffie and M. Hellman, *New directions in cryptography*, IEEE Trans. Inform. Theory **22** (1976), 644–654.

[8] J. Dj. Golić and M. Živković, *On the linear complexity of nonuniformly decimated PN-sequences*, IEEE Trans. Inform. Theory **34** (1988), 1077–1079.

[9] J. Dj. Golić, *On the linear complexity of functions of periodic GF(q) sequences*, IEEE Trans. Inform. Theory **35** (1989), 69–75.

[10] J. Dj. Golić and M. Mihaljević, *Minimal linear equivalent analysis of a variable-memory binary sequence generator*, IEEE Trans. Inform. Theory **36** (1990), 190–192.

[11] J. Dj. Golić and M. Mihaljević, *A noisy clock-controlled shift register cryptanalysis concept based on sequence comparison approach*, *Advances in Cryptology – EUROCRYPT '90* (I. B. Damgard, ed., Springer-Verlag), Lecture Notes in Computer Science **473** (1991), 487–491.

[12] J. Dj. Golić, *The number of output sequences of a binary sequence generator*, *Advances in Cryptology - EUROCRYPT '91* (D. V. Davies, ed., Springer-Verlag), Lecture Notes in Computer Science **547** (1991), 160–167.

[13] J. Dj. Golić and M. Mihaljević, *A generalized correlation attack on a class of stream ciphers based on the Levenshtein distance*, J. Cryptology **3** (1991), 201–212.

[14] J. Dj. Golić and S. Petrović, *Constrained edit distance for a memoryless function of strings*, invited introductory paper, *Proc. 2. Spanish Conference on Cryptology*, Madrid, Spain, 1992, 1–23.

[15] J. Dj. Golić and S. Petrović, *A generalized correlation attack with a probabilistic constrained edit distance*, *Advances in Cryptology – EUROCRYPT '92* (R. A. Rueppel, ed., Springer-Verlag), Lecture Notes in Computer Science **658** (1993), 472–476.

[16] J. Dj. Golić, *Correlation via linear sequential circuit approximation of combiners with memory*, *Advances in Cryptology – EUROCRYPT '92* (R. A. Rueppel, ed., Springer-Verlag), Lecture Notes in Computer Science **658** (1993), 113–123.

[17] J. Dj. Golić, *On the security of shift register based keystream generators*, *Fast Software Encryption – Cambridge '93* (R. Anderson, ed., Springer-Verlag), Lecture Notes in Computer Science **809** (1994), 90–100.

[18] J. Dj. Golić, M. Salmasizadeh, and E. Dawson, *Autocorrelation weakness of multiplexed sequences*, *Proc. 1994 International Symposium on Information Theory and Its Applications*, Sydney, Australia, 1994, 983–987.

[19] J. Dj. Golić, *Noiseless coding for multiple channels*, *Proc. 1994 International Symposium on Information Theory and Its Applications*, Sydney, Australia, 1994, 787–792.

[20] J. Dj. Golić, *Intrinsic statistical weakness of keystream generators*, *Advances in Cryptology - ASIACRYPT '94* (J. Pieprzyk, R. Safavi-Naini, eds., Springer-Verlag), Lecture Notes in Computer Science **917** (1995), 91–103.

[21] J. Dj. Golić, *Linear cryptanalysis of stream ciphers*, *Fast Software Encryption – Leuven '94* (B. Preneel, ed., Springer-Verlag), Lecture Notes in Computer Science **1008** (1995), 169.

[22] J. Dj. Golić and L. O'Connor, *Embedding and probabilistic correlation attacks on clock-controlled shift registers*, *Advances in Cryptology – EUROCRYPT '94* (A. De Santis, ed., Springer-Verlag), Lecture Notes in Computer Science **950** (1995), 230–243.

[23] J. Dj. Golić, *Towards fast correlation attacks on irregularly clocked shift registers*, *Advances in Cryptology – EUROCRYPT '95* (L. C. Guillou, J.-J. Quisquater, eds., Springer-Verlag), Lecture Notes in Computer Science **921** (1995), 248–262.

[24] J. Dj. Golić and S. Petrović, *Constrained many-to-one string editing with memory*, Information Sciences **86** (1995), 61–76.

[25] J. Dj. Golić, *On decimation of linear recurring sequences*, The Fibonacci Quarterly **33** (1995), 407–411.

[26] J. Dj. Golić, M. Salmasizadeh, A. Clark, A. Khodkar, and E. Dawson, *Discrete optimisation and fast correlation attacks*, *Cryptography: Policy and Algorithms – Brisbane '95* (E. Dawson, J. Golić, eds., Springer-Verlag), Lecture Notes in Computer Science **1029** (1996), 186–200.

[27] J. Dj. Golić and L. O'Connor, *A cryptanalysis of clock-controlled shift registers with multiple steps*, *Cryptography: Policy and Algorithms – Brisbane '95* (E. Dawson, J. Golić, eds., Springer-Verlag), Lecture Notes in Computer Science **1029** (1996), 174–185.

[28] J. Dj. Golić, *On the security of nonlinear filter generators*, *Fast Software Encryption – Cambridge '96* (D. Gollmann, ed., Springer-Verlag), Lecture Notes in Computer Science **1039** (1996), 173–188.

[29] J. Dj. Golić, *Linear models for keystream generators*, IEEE Trans. Comput. **45** (1996), 41–49.

[30] J. Dj. Golić, *Correlation properties of a general binary combiner with memory*, J. Cryptology **9** (1996), 111–126.

[31] J. Dj. Golić, *Fast low order approximation of cryptographic functions*, *Advances in Cryptology – EUROCRYPT '96* (U. Maurer, ed., Springer-Verlag), Lecture Notes in Computer Science **1070** (1996), 268–282.

[32] J. Dj. Golić and S. Petrović, *Correlation attacks on clock-controlled shift registers in keystream generators*, IEEE Trans. Comput. **45** (1996), 482–486.

[33] J. Dj. Golić, *Constrained embedding probability for two binary strings*, SIAM Journal on Discrete Mathematics **9** (1996), 360–364.

[34] J. Dj. Golić, *Computation of low-weight parity-check polynomials*, Electronics Letters **32** (1996), 1981–1982.

[35] J. Dj. Golić, *Conditional correlation attack on combiners with memory*, Electronics Letters **32** (1996), 2193–2195.

[36] J. Dj. Golić, *On period of multiplexed sequences*, *Information Security and Privacy – Wollongong '96* (J. Pieprzyk, J. Seberry, eds., Springer-Verlag), Lecture Notes in Computer Science **1172** (1996), 158–168.

[37] J. Dj. Golić, *Edit distance correlation attacks on clock-controlled combiners with memory*, *Information Security and Privacy – Wollongong '96* (J. Pieprzyk, J. Seberry, eds., Springer-Verlag), Lecture Notes in Computer Science **1172** (1996), 169–181.

[38] J. Dj. Golić, *Cryptanalysis of alleged A5 stream cipher*, *Advances in Cryptology – EUROCRYPT '97* (W. Fumy, ed., Springer-Verlag), Lecture Notes in Computer Science **1233** (1997), 239–255.

[39] J. Dj. Golić, *Linear statistical weakness of alleged RC4 keystream generator*, *Advances in Cryptology – EUROCRYPT '97* (W. Fumy, ed., Springer-Verlag), Lecture Notes in Computer Science **1233** (1997), 226–238.

[40] J. Dj. Golić and R. Menicocci, *Edit distance correlation attack on the alternating step generator*, *Advances in Cryptology – CRYPTO '97* (B. Kaliski, ed., Springer-Verlag), Lecture Notes in Computer Science **1294** (1997), 499–512.

[41] J. Dj. Golić, M. Salmasizadeh, L. Simpson, and E. Dawson, *Fast correlation attacks on nonlinear filter generators*, Information Processing Letters **64** (1997), 37–42.

[42] J. Dj. Golić, *Universal stream ciphers*, *Cryptography*, *Dagstuhl-Seminar-Report*, vol. 190, by invitation, Dagstuhl, Germany, 1997, p. 5.

[43] J. Dj. Golić, *On matroid characterization of ideal secret sharing schemes*, J. Cryptology **11** (1998), 75–86.

[44] J. Dj. Golić, *Periods of interleaved and nonuniformly decimated sequences*, IEEE Trans. Inform. Theory **44** (1998), 1257–1260.

[45] J. Dj. Golić, M. Salmasizadeh, and E. Dawson, *Fast correlation attacks on the summation generator*, J. Cryptology, to appear.

[46] J. Dj. Golić, *Modes of use of stream ciphers*, submitted.

[47] J. Dj. Golić, A. Clark, and E. Dawson, *Generalized inversion attack on nonlinear filter generators*, submitted.

[48] J. Dj. Golić, *Linear models for a time-variant permutation generator*, submitted.

[9] J. Dj. Golić, M. Salmasizadeh, and E. Dawson, *Statistical weakness of multiplexed sequences*, submitted.

[50] J. Dj. Golić and R. Menicocci, *Edit probability correlation attack on the alternating step generator*, submitted.

[51] J. Dj. Golić, *Edit distances and probabilities for correlation attacks on irregularly clocked combiners*, submitted.

[52] J. Dj. Golić, *A cryptanalysis of three mutually clock-controlled stop/go shift registers*, submitted.

[53] D. Gollmann and W. G. Chambers, *Clock-controlled shift registers: A review*, IEEE J. Select. Areas Commun. **7** (1989), 525–533.

[54] San Francisco, Holden-Day, 1967.

[55] H. Gustafson, E. Dawson, and J. Dj. Golić, *Randomness measures related to subset occurrence*, *Cryptography: Policy and Algorithms – Brisbane '95* (E. Dawson, J. Golić, eds., Springer-Verlag), Lecture Notes in Computer Science **1029** (1996), 132–143.

[56] H. Gustafson, E. Dawson, and J. Dj. Golić, *Automated statistical methods for measuring the strength of block ciphers*, Statistics and Computing **7** (1997), 125–135.

[57] H. Gustafson, E. Dawson, J. Dj. Golić, and A. Pettitt, *Methods for testing subblock patterns*, submitted.

[58] A. Klapper, *The vulnerability of geometric sequences based on fields of odd characteristic*, J. Cryptology **7** (1994), 33–51.

[59] A. Klapper and M. Goresky, *Feedback shift registers,* 2*-adic span, and combiners with memory*, J. Cryptology **10** (1997), 111–147.

[60] S.-J. Lee, S.-J. Park, and S.-C. Goh, *On the security of the Gollmann cascades*, *Advances in Cryptology – CRYPTO '95* (D. Coppersmith, ed., Springer-Verlag), Lecture Notes in Computer Science **963** (1995), 148–157.

[61] *A universal statistical test for random bit generators*, J. Cryptology **5** (1992), 89–105.

[62] W. Meier and O. Staffelbach, *Fast correlation attacks on certain stream ciphers*, J. Cryptology **1** (1989), 159–176.

[63] W. Meier and O. Staffelbach, *Nonlinearity criteria for cryptographic functions*, *Advances in Cryptology – EUROCRYPT '89* (J.-J. Quisquater, J. Vandewalle, eds., Springer-Verlag), Lecture Notes in Computer Science **434** (1990), 549–562.

[64] W. Meier and O. Staffelbach, *Correlation properties of combiners with memory in stream ciphers*, J. Cryptology **5** (1992), 67–86.

[65] A. Menezes, P. van Oorschot, and S. Vanstone, *Handbook of Applied Cryptography*, CRC Press, New York, 1997.

[66] R. Menicocci, *Cryptanalysis of a two-stage Gollmann cascade generator*, *Proc. SPRC '93*, Rome, Italy, 1993, 62–69.

[67] R. Menicocci and J. Dj. Golić, *Correlation attacks on stop/go and up/down cascades*, IEEE Trans. Inform. Theory, to appear.

[68] M. Mihaljević and J. Dj. Golić, *A fast iterative algorithm for a shift register initial state reconstruction given the noisy output sequence*, *Advances in Cryptology – AUSCRYPT '90* (J.

Seberry, J. Pieprzyk, eds., Springer-Verlag), Lecture Notes in Computer Science **453** (1990), 165–175.

[69] M. Mihaljević and J. Dj. Golić, *A comparison of cryptanalytic principles based on iterative error-correction*, *Advances in Cryptology – EUROCRYPT '91* (D. V. Davies, ed., Springer-Verlag), Lecture Notes in Computer Science **547** (1991), 527–531.

[70] M. Mihaljević and J. Dj. Golić, *Convergence of a Bayesian iterative error-correction procedure on a noisy shift register sequence*, *Advances in Cryptology – EUROCRYPT '92* (R. A. Rueppel, ed., Springer-Verlag), Lecture Notes in Computer Science **658** (1993), 124–137.

[71] S. Petrović and J. Dj. Golić, *String editing under a combination of constraints*, Information Sciences **74** (1993), 151–163.

[72] R. A. Rueppel, *Analysis and Design of Stream Ciphers*, Springer-Verlag, Berlin, 1986.

[73] R. A. Rueppel, *Stream ciphers*, *Contemporary Cryptology: The Science of Information Integrity* (G. J. Simmons, ed.), IEEE Press, New York, 1992, 65–134.

[74] B. Schneier, *Applied Cryptography*, Wiley, New York, 1996.

[75] P. D. Seymour, *On secret-sharing matroids*, J. Combin. Theory Ser. B **56** (1992), 69–73.

[76] A. Shamir, *How to share a secret*, Comm. ACM **22** (1979), 612–613.

[77] *Communication theory of secrecy systems*, Bell System Technical Journal **28** (1949), 656–715.

[78] T. Siegenthaler, *Correlation immunity of nonlinear combining functions for cryptographic applications*, IEEE Trans. Inform. Theory **30** (1984), 776–780.

[79] T. Siegenthaler, *Decrypting a class of stream ciphers using ciphertext only*, IEEE Trans. Comput. **34** (1985), 81–85.

[80] L. Simpson, J. Dj. Golić, and E. Dawson, *A probabilistic correlation attack on the shrinking generator*, *Information Security and Privacy – Brisbane '98* (C. Boyd, E. Dawson, eds., Springer-Verlag), Lecture Notes in Computer Science **1438** (1998), 147–158.

[81] D. J. A. Welsh, *Matroid Theory*, Academic Press, London, 1976.

[82] K. C. Zeng, C. H. Yang, and T. R. N. Rao, *On the linear consistency test (LCT) in cryptanalysis and its applications*, *Advances in Cryptology – CRYPTO '89* (G. Brassard, ed., Springer-Verlag), Lecture Notes in Computer Science **435** (1990), 164–174.

[83] M. Živković, *On two probabilistic decoding algorithms for binary linear codes*, IEEE Trans. Inform. Theory **37** (1991), 1707–1716.

[84] M. Živković, *An algorithm for the initial state reconstruction of the clock-controlled shift register*, IEEE Trans. Inform. Theory **37** (1991), 1488–1490.

Elektrotehnički fakultet
Bulevar Revolucije 73
11001 Beograd
Yugoslavia
golic@galeb.etf.bg.ac.yu

(Received 01 06 1998)
(Revised 24 11 1998)