Università degli Studi Roma Tre

Facoltà di Scienze Matematiche Fisiche e Naturali

Synthesis

Stream Ciphers: from Correlation Attacks to the Cube Attack

Candidate

Andrea Agnesse

matricola: 269686

Supervisor

prof. Marco Pedicini

Academic year: 2007/2008

MSC AMS: 06E30; 11T71; 68P25; 94A55; 94A60

keywords: stream ciphers, LFSR keystream generator, linear recurring sequences, combining functions, correlation attack, fast correlation attack, cube attack

Synthesis

For centuries men have had the necessity of communicate secretly. The first attempt used to reach this result has been *steganography*, which consists in simply hide the message; the world come from the Greek and means "concealed writing". The first examples of steganography are mentioned by Herodotus, around 440 B.C., in his work "The Histories": Demaratus, who sent a warning about a forthcoming attack to Greece by writing it directly on the wooden backing of a wax tablet before applying its beeswax surface and Histiaeus, who sand his most trusted slave after having shaved his head and tattooed a message on it.

The most common and developed set of techniques used to let two parties secretly communicate is, however, *cryptography*, which means "secret writing". Differently from steganography, cryptography is the science that studies ways to hide information, without necessarily hide the message, so that it can not be understood even if an opposite part find it.

The model used in cryptography is the following: two parts, Alice and Bob, want to communicate without let Freddie[1] read their messages. To achieve their aim, Alice and Bob send themself encrypted messages that only the other one can read, while Freddie's aim is to menage to understand the messages, i.e. to find a way to break the encryption used by Alice and Bob.

To formalize this process we use the following definitions.

Definition Let $\mathcal{A}$ be an alphabet of q symbols.

A *plaintext* is a sequence of n symbols from $\mathcal{A}$, $\boldsymbol{m} = m_1 m_2 \ldots m_n \in \mathcal{A}^n$, while the *ciphertext* is the result of the encryption transformation, and is

[1] Mr. Bad Guy.

usually a sequence of n symbols from the same set $\mathcal{A}$, $\boldsymbol{c} = c_1c_2\ldots c_n \in \mathcal{A}^n$. We denote the set of all the plaintexts as $\mathcal{M}$, while $\mathcal{C}$ is the set of all the ciphertexts. We denote by $\mathcal{K}$ the *keyspace*, i.e. the set of all the keys.

An *encryption* is a function $E_k : \mathcal{M} \times \mathcal{K} \to \mathcal{C}$ that receives a plaintext $\boldsymbol{m}$ and produces a ciphertext $\boldsymbol{c}$, according to $k \in \mathcal{K}$. The inverse transformation $D_{k'} : \mathcal{C} \times \mathcal{K} \to \mathcal{M}$ is called *decryption*.

A *cipher* is a pair of functions (E, D) such that for any plaintext $\boldsymbol{m} \in \mathcal{M}$ we have $D_{k'}\left(E_k(\boldsymbol{m})\right) = \boldsymbol{m}$, where k' is the decryption key corresponding to the key k used for the encryption.

If the two keys used are equals, $k = k'$, or one can be easily drawn from the other, we talk about *symmetric encryption*. If so, the scheme of the communication between Alice and Bob with the attack of Freddie is the one illustrated in Fig. 1.

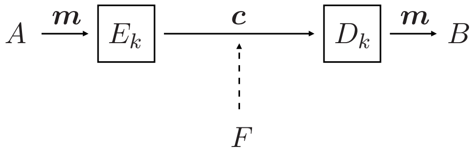

Fig. 1: Scheme of a symmetric encryption; $\boldsymbol{c}$ is the encrypted text.

A basic require for the cipher to be safe is that the keyspace $\mathcal{K}$ is large enough to do not permit a *brute force attack*, that is an exhaustive search of the only right key by trying all of them.

Definition An *attack* is a strategy that contributes to reduce significantly the set of possible keys.

Note that the key is the most important part of all the system. As stated by the “Kerckhoffs’ principle”, in fact, the cipher has to be safe even if the attacker have a detailed description of it (but lacks the key). This revolutionary concept was formulated in 1883 and restated by Claude Shannon, the inventor of information theory and the fundamentals of theoretical cryptography, as “the knowledge of the cipher should not drop any information about the message”.

One the first (and most famous) example of symmetric cryptographic scheme is *Caesar cipher*, in which every letter of the message is substituted with the one three positions further down in the alphabet (see Table 1).

plaintext digit	A	B	C	D	...	T	U	V	Z
ciphertext digit	D	E	F	G	...	Z	A	B	C

Table 1: Substitutions in the Caesar cipher.

Note that this cipher is absolutely weak, as the keyspace consists, at most, in 26 keys, so that an exhaustive search of the only right key is possible.

With the development of the techniques of cryptography, various physical devices and aids have become to arise, to help people do encryption faster, or because the key was a physical characteristic of the devise itself. One of the most famous ancient of such devises was the *scytale*, which consists in a cylinder (such as a stick) with a strip of leather wound around it on which the message is written longwise (see Fig. 2). In this case the key is represented by the diameter of the cylinder used to encrypt the message, since to decrypt the message a cylinder with the same thickness is needed.

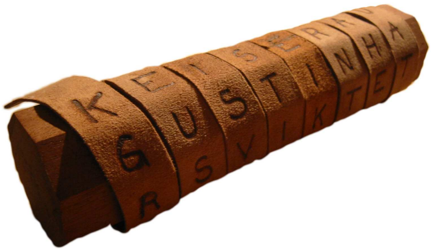

Fig. 2: A scytale; image from Wikipedia.

In the early of 20th century several mechanical encryption/decryption devices has been invented invented to perform faster and more complex ciphers. An examples of such mechanical devices are the *rotor machines* used, for instance, in the *Enigma machine* during World War II by German military (see Fig. 3).

Fig. 3: Enigma; image from Wikipedia.

The development of more complex ciphers was made possible by the development of digital computers, which, moreover, let encrypt every type of data representable within the computer memory, such as pictures, videos and databases.

A major advantage of the use of computers in cryptography was the improvement of the computational power, which lead, in the 70s, to the development of *public key cryptography*, in which the key k' used to decrypt is different from the one used to encrypt k. In that way the key k (called *public key*) can be distributed to anyone who wants to communicate with us, since only the *secret key* k' can decrypt the message. Public key cryptography scheme are usually based on difficult mathematical problems, such as the discrete logarithms; they, however, still does not give the same good performance of symmetric key encryption, so they are used above all to encrypt small messages, such as symmetric keys.

A basic classification of symmetric key encryption is the one between *block ciphers* and *stream ciphers*. Briefly, we can distinguish them saying that a block chiper splits the message in different parts of the same length, and then it encrypts each part separately. On the other hand, stream ciphers operate on a smaller number of symbols, most of times on a single digit, performing a different encryption for each one. The sequence of keys the stream cipher uses to make the various different encryptions is called *keystream*.

To reach the *perfect security* the keystream should be of the same length of the message, should be used only once and should be chosen completely random. This last condition is the most problematic in practise, as there should be found a way to secretly communicates the key between the parts. For that reason stream ciphers are usually used with a *keystream generator*[2], which is an algorithm (implemented in hardware or software) used to generate a deterministic sequence of keys, but the most possible random-looking.
To produce such a sequences each part of the keystream generator must be chosen very carefully, as there is the risk that the keystream generated can appear in some sense predictable, thus not giving a strong encryption.

In this thesis we present three innovative attacks against combining LFSR keystream generators, which are used in many areas for their high speed and efficiency and low computational resources required. All of these attacks explore, in some sense, a weakness which was unknown when they have been invented, giving, in that way new criteria designs of such keystream generators.

In Chapter 1 we give the basic mathematical definitions and introduce some properties of the algebraic structure used in the rest of this thesis. We give also an overview on probability and information theory.

[2] Also called *pseudo-noise generator*.

In Chapter 2 we describe and specifies the cryptographic properties of each part of LFSR keystream generators, which scheme is illustrated in Fig 4.

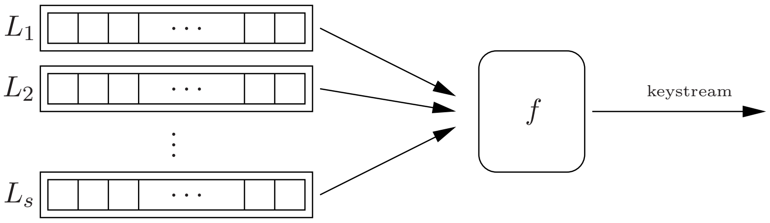

Fig. 4: Scheme of a LFSR keystream generator: on the left there are the registers, while f is the combining function.

A linear feedback shift register is a register made of several stages, each one capable of storing a bit (or, more generally, an element from a finite field $\mathbb{F}_q$). All the stages are synchronized by an external clock: at each unit time each register releases its value and accepts another one. As the stages are concatenated together, the internal states of the register actually shifts of one position, and a bit of output is generated, while on the opposite side, the last stage accepts a linear combination of the values of the stages.

In Section 2.1 we studied the properties of the sequences generated by such a register. We also show how these properties depend on the initial state and (above all) the feedback connection that generates the linear combination of the stages.

One of the way of designing a keystream generator based on LFSR is through the use of a boolean function, which, at each unit time, combines the outputs of several LFSR and produces a single digit of keystream. In Section 2.2 we study the cryptographic properties of these functions.

In Chapter 3 we present the Correlation Attack by Siegenthaler [44], which, in 1984/1985, had the idea of using a statistical weakness of a not properly chosen combining function in LFSR keystream generator (see

Fig. 5). In fact, Siegenthaler showed that if the function makes possible a correlation between the output of some registers and the ciphertext, obtained by a bitwise XOR between the plaintext message and the keystream, recovering the initial state and the feedback connections of the register become possible in a ciphertext-only attack.

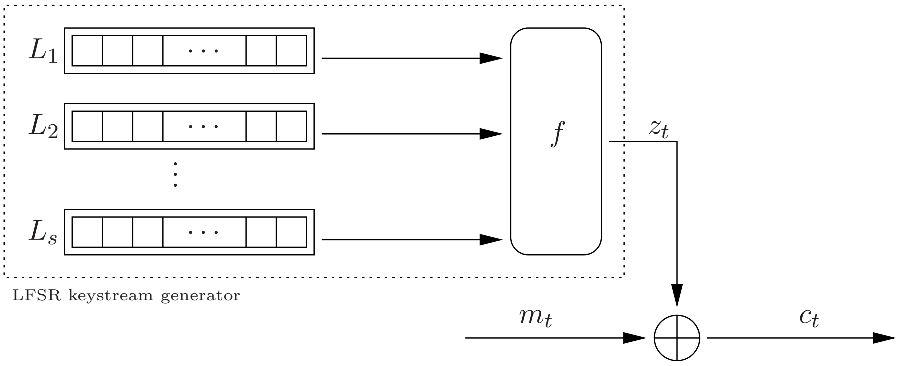

Fig. 5: Model used by Siegenthaler in his Correlation Attack.

The strategy he set out was to perform a some kind of brute force attack on all the pairs feedback connection/initial state of each register, separately from all the other ones. In this way the attacker can reduce the complexity of a brute force attack from $\prod_{i=1}^{s} K_i$ to $\sum_{i=1}^{s} K_i$, where K_i is the total number of possible combination for the ith register.

Despite the great limit of this attack, i.e. the length of the register, since if ti is too long the exhaustive search is infeasible, Siegenthaler showed in practice a new weakness of most LFSR keystream generator proposed at that time. He also stressed the importance of a new cryptographic property, the *correlation immunity*, for the combining functions.

In Chapter 4 we present an improvements of the attack presented in the previous chapter, the Fast Correlation Attack developed by Meier and Staffelbach in 1988/1989. The aim of their two algorithms was to recover the output of a single register (whose feedback connections are supposed to be known) from the observed keystream.

The model used in their attacks is the one illustrated in Fig. 6: the output of the register is thought to be filtered by a binary symmetric memoryless source, which summarizes the action of the other registers and of the combining function over the output itself. In this scenario, the problem of finding the only right output of the register is equivalent to the one of decoding a message transitted through a noisy channel, where the correlation became the probability of each bit of not being flipped. The two algorithms ideated by Meier and Staffelbach used the fact that the output of the register has a known structure, given by the linear relation it satisfies, which can be used to correctly decoding it from the noisy keystream.

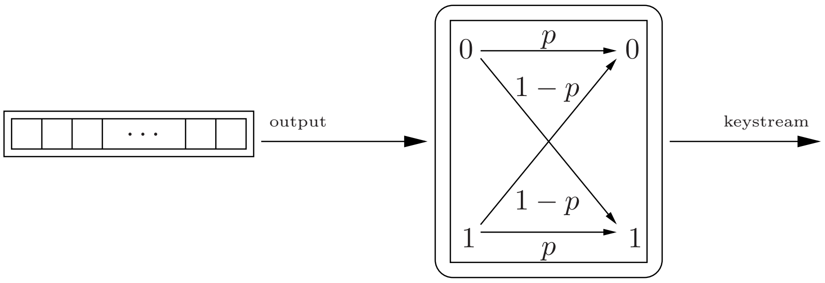

Fig. 6: Scheme of a LFSR keystream generator: on the left there are the registers, while f is the combining function.

These attacks suffered of a very important limitation, since they are successfully only if the number of stages actually used in the feedback linear combination of the register is rather small (almost less than 10). Their ideas were, however, very innovative, as they showed how a cryptographic problem could be seen as a decoding problem, and how the techniques used in this area could be used do recover the initial state of a LFSR. This led to numerous improvements of their attacks, since others began to apply different techniques of decoding, managing to overcome the limit of the few feedback connections required to perform the attack.

In Chapter 5 we describe the new attack ideated by Shamir and his studend Dinur at the end of 2008, the Cube Attack. This chosen-plaintext attack has very few innovations, since, as stated in their paper by their author, their work has been limited to collect and re-organize a series of techniques already known, producing, however, an original new attack suitable for block and stream ciphers. This fact has opened a dispute with Michael Vielhaber, the inventor of AIDA[3], a previous attack directed to the stream cipher *Trivium*, which directly accused Shamir and Dinur of plagiarism.

The Cube Attack is based on the possibility to describe the cipher as a multivariate polynomial p over $\mathbb{F}_2$ in $m + n$ variables: the first ones are called *public variables*, and stores the IV bits, while the latters are called *secret variables* and store the key bits. The aim of this attack is to recover the n key bits used in the cipher. to do that Shamir and Dinur proved two therems, which are the basis of the attack. They, in fact, permit to obtain, under certain conditions, a linear system in only the secret variables and of which the free terms can be easily calculated in the phase of the attack, so that it can be easily solved in polynomial time.

The most onerous part of the attack is, however, the preprocessing phase, in which the attacker tries to find enough monomials t_I made of only public variables; for each such monomial the polynomial p can be written as $p = t_I p_{S(I)} + q_I$, where $p_{S(I)}$ is called the *superpoly* of the term t_I in p. Due to the fact that $x^2 = x$ for all $x \in \mathbb{F}_2$, we can suppose that all the variables appear with the exponent less than 2 in every monomial of p, so in every term t_I and its superpoly $p_{S(I)}$ always different variables appear. The aim of

[3] Algebraic IV Differential Attack.

this phase is to choose public terms for which the corresponding superpoly is linear and non constant, so that we can set out a linear system of secret variables with a chosen-plaintext attack.

The unsolved question is how to efficiently find enough public *maxterms*, i.e. terms such that their superpolies contains only secret variables. Note that the question of how to find some linear (in the secret vars) parts of the full ANF (algebraic normal form) in some polynomial-bounded time remains unsolved also in the AIDA paper by Vielhaber, fact that contributes to his accuses, as the Cube Attack does not resolve that fundamental point.

In Section 5.4 we present an original discussion oriented on finding similar results of the two theorems in the paper by Shamir and Dinur, but in a generic finite field $\mathbb{F}_q$. In our discussion we show that some of the properties are not valid, however we are able to conduct the attack.

In Appendix A we present the source code of the implementations we realized of the attacks presented in Chapters 3, 4 and 5. All the implementations are realized with Wolfram Mathematica, and are suitable to be tested with any finite field $\mathbb{F}_q$, having used the package `FiniteFields`, already present in the Mathematica installation.

In Appendix B we present the *Berlekamp-Massey Algorithm*, which is used to find the shortest LFSR that generates a given output of elements in a finite field $\mathbb{F}_q$ in polynomial time, given an upper bound of its length. The length of such a LFSR is called *linear complexity* of the sequence. This algorithm is at the basis of various attacks, which can performed if the combining function of a LFSR keystream generator is linear, since the resulted keystream has a too short linear complexity, and so it can described by a set of linear equations that can be easily solved.

References

[1] Chu-Wee Lim Aileen Zhang and Khoongming Khoo. Extensions of the Cube Attack. Cryptology ePrint Archive, Report 2009/049, 2009. `http://eprint.iacr.org/`.

[2] Paul C. van Oorschot Alfred J. Menezes and Scott A. Vanstone. *Handbook of Applied Cryptography*. CRC Press, fifth printing edition, 2001. `http://www.cacr.math.uwaterloo.ca/hac/`.

[3] Nadia Ben Atti, Gema M. Diaz–Toca, and Henri Lombardi. The Berlekamp-Massey Algorithm revisited. *Appl. Algebra Eng., Commun. Comput.*, 17(1):75–82, 2006.

[4] Elad Barkan and Eli Biham. Conditional estimators: An effective attack on A5/1. In *Selected Areas in Cryptography*, pages 1–19, 2005.

[5] S S Bedi and N Rajesh Pillai. Cube Attacks on Trivium. Cryptology ePrint Archive, Report 2009/015, 2009. `http://eprint.iacr.org/`.

[6] J.O. Brüer. On nonlinear combinations of linear shift register sequences. *Proc. IEEE ISIT*, 1982.

[7] P. Camion, C. Carlet, P. Charpin, and N. Sendrier. On correlation-immune functions. In *Advances in cryptology—CRYPTO '91 (Santa Barbara, CA, 1991)*, volume 576 of *Lecture Notes in Comput. Sci.*, pages 86–100. Springer, Berlin, 1992.

[8] A. Canteaut. Fast correlation attacks against stream ciphers and related open problems. pages 49–54, Oct. 2005.

[9] Anne Canteaut and Eric Filiol. Ciphertext only reconstruction of stream ciphers based on combination generators. In *FSE*, pages 165–180, 2000.

[10] Anne Canteaut and Michaël Trabbia. Improved fast correlation attacks using parity-check equations of weight 4 and 5. In *EUROCRYPT*, pages 573–588, 2000.

[11] Seongtaek Chee, Sangjin Lee, Kwangjo Kim, and Daeho Kim. Correlation immune functions with controllable nonlinearity. *ETRI Journal*, 19(4):389–401, December 1997.

[12] Vladimor V. Chepyzhov, Thomas Johansson, and Ben Smeets. A simple algorithm for fast correlation attacks on stream ciphers. In *FSE '00: Proceedings of the 7th International Workshop on Fast Software Encryption*, pages 181–195, London, UK, 2001. Springer-Verlag.

[13] Vladimor V. Chepyzhov and Ben J. M. Smeets. On a fast correlation attack on certain stream ciphers. In *EUROCRYPT*, pages 176–185, 1991.

[14] Itai Dinur and Adi Shamir. Cube Attacks on Tweakable Black Box Polynomials. Cryptology ePrint Archive, Report 2008/385, 2008. `http://eprint.iacr.org/`.

[15] Itai Dinur and Adi Shamir. Side Channel Cube Attacks on Block Ciphers. Cryptology ePrint Archive, Report 2009/127, 2009. `http://eprint.iacr.org/`.

[16] Patrik Ekdahl and Thomas Johansson. Another attack on A5/1. *IEEE Transactions on Information Theory*, 49(1):284–289, 2003.

[17] Eric Filiol and Caroline Fontaine. Highly nonlinear balanced Boolean functions with a good correlation-immunity. In *Advances in cryptology—EUROCRYPT '98 (Espoo)*, volume 1403 of *Lecture Notes in Comput. Sci.*, pages 475–488. Springer, Berlin, 1998.

[18] R. Gallager. Low-density parity-check codes. *Information Theory, IRE Transactions on*, 8(1):21–28, January 1962.

[19] Solomon W. Golomb and Guang Gong. *Signal design for good correlation.* Cambridge University Press, Cambridge, 2005. For wireless communication, cryptography, and radar.

[20] K. Gopalakrishnan and Douglas R. Stinson. Three characterizations of non-binary correlation-immune and resilient functions. *Des. Codes Cryptography*, 5(3):241–251, 1995.

[21] Martin Hell, Thomas Johansson, and Willi Meier. Grain - a stream cipher for constrained environments. Technical report, 2005/010, ECRYPT (European Network of Excellence for Cryptology, 2005.

[22] Thomas Johansson and Fredrik Jönsson. Fast correlation attacks based on turbo code techniques. In *CRYPTO '99: Proceedings of the 19th Annual International Cryptology Conference on Advances in Cryptology*, pages 181–197, London, UK, 1999. Springer-Verlag.

[23] Thomas Johansson and Fredrik Jönsson. Improved fast correlation attacks on stream ciphers via convolutional codes. In *EUROCRYPT*, pages 347–362, 1999.

[24] Thomas Johansson, Fredrik Jönsson, and Student Member. Theoretical analysis of a correlation attack based on convolutional codes. In *In Ezio*

Biglieri and Sergio Verdu, editors, IEEE International Symposium on Information Theory 2000, pages 21–2, 2000.

[25] Edwin L. Key. An analysis of the structure and complexity of nonlinear binary sequence generators. *IEEE Trans. Inform. Theory*, 22(6):732–736, 1976.

[26] Rudolf Lidl and Harald Niederreiter. *Introduction to finite fields and their applications*. Cambridge University Press, Cambridge, revised edition, 1994.

[27] Mulan Liu, Peizhong Lu, and G.L. Mullen. Correlation-immune functions over finite fields. *Information Theory, IEEE Transactions on*, 44(3):1273–1276, May 1998.

[28] David MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[29] J. Massey. Shift-register synthesis and BCH decoding. *Information Theory, IEEE Transactions on*, 15(1):122–127, Jan 1969.

[30] Alexander Maximov, Thomas Johansson, and Steve Babbage. An improved correlation attack on A5/1. In *Selected Areas in Cryptography*, pages 1–18, 2004.

[31] Willi Meier and Othmar Staffelbach. Fast correltaion attacks on stream ciphers (extended abstract). In *EUROCRYPT*, pages 301–314, 1988.

[32] Willi Meier and Othmar Staffelbach. Fast correlation attacks on certain stream ciphers. *J. Cryptology*, 1(3):159–176, 1989.

[33] Miodrag J. Mihaljevic and Jovan D. Golić. A fast iterative algorithm for a shift register initial state reconstruction given the noisy output

sequence. In *AUSCRYPT '90: Proceedings of the international conference on cryptology on Advances in cryptology*, pages 165–175, New York, NY, USA, 1990. Springer-Verlag New York, Inc.

[34] James S. Milne. Fields and Galois Theory (v4.21), 2008. Available at `www.jmilne.org/math/`.

[35] Geffe P. How to protect data with ciphers that are really hard to break. *Electronics*, (4):99–101, January 1973.

[36] Walter T. Penzhorn. Correlation attacks on stream ciphers: Computing low-weight parity checks based on error-correcting codes. In *FSE*, pages 159–172, 1996.

[37] V.S. Pless. Encryption schemes for computer confidentiality. *IEEE Transactions on Computers*, 26(11):1133–1136, 1977.

[38] Palash Sarkar. A note on the spectral characterization of correlation immune boolean functions. *Inf. Process. Lett.*, 74(5-6):191–195, 2000.

[39] Palash Sarkar and Subhamoy Maitra. Balancedness and correlation immunity of symmetric boolean functions. *Electronic Notes in Discrete Mathematics*, 15:176–181, 2003.

[40] Bruce Schneier. *Applied cryptography (2nd ed.): protocols, algorithms, and source code in C.* John Wiley & Sons, Inc., New York, NY, USA, 1995.

[41] Jennifer Seberry, Xian-Mo Zhang, and Yuliang Zheng. On constructions and nonlinearity of correlation immune functions. *Lecture Notes in Computer Science*, 765:181–197, 1994.

[42] Claude E. Shannon. *A Mathematical Theory of Communication.* CSLI Publications, 1948.
http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html.

[43] T. Siegenthaler. Correlation-immunity of nonlinear combining functions for cryptographic applications. *IEEE Trans. Inform. Theory*, 30(5):776–780, 1984.

[44] T. Siegenthaler. Decrypting a class of stream ciphers using ciphertext only. *IEEE Trans. Computers*, 34(1):81–85, 1985.

[45] Michael Vielhaber. Breaking ONE.FIVIUM by AIDA an Algebraic IV Differential Attack. Cryptology ePrint Archive, Report 2007/413, 2007.
http://eprint.iacr.org/.

[46] Guo Zhen Xiao and James L. Massey. A spectral characterization of correlation-immune combining functions. *IEEE Trans. Inform. Theory*, 34(3):569–571, 1988.

[47] Kencheng Zeng and Minqiang Huang. On the linear syndrome method in cryptoanalysis. In *CRYPTO '88: Proceedings of the 8th Annual International Cryptology Conference on Advances in Cryptology*, pages 469–478, London, UK, 1990. Springer-Verlag.

[48] Kencheng Zeng, Chung-Huang Yang, and T. R. N. Rao. An improved linear syndrome algorithm in cryptanalysis with applications. In *CRYPTO '90: Proceedings of the 10th Annual International Cryptology Conference on Advances in Cryptology*, pages 34–47, London, UK, 1991. Springer-Verlag.