DESIGN OF COMBINERS TO PREVENT DIVIDE AND CONQUER ATTACKS

T. Siegenthaler
Institute for Communication Technology
Federal Institute of Technology
8092 Zurich, Switzerland

Abstract

A finite state machine driven by n independent sources each generating a q-ary sequence is investigated. The q-ary output sequence of that device is considered as the running-key sequence in a stream cipher. Possible definitions for Correlation-Immunity are discussed and a simple condition is given which ensures that divide-and-conquer attacks on such generators are prevented.

I Introduction

A common form of running key generators for use in stream ciphers consists of n driving sources and some combiner. We assume in this section that each of the sources independently generates a sequence of q-ary random variables and that a finite state machine (FSM) combines the n input sequences $x_{1,j}, x_{2,j}, \ldots, x_{n,j}$ to an output sequence z_j, $j=0,1,\ldots$. A FSM is a system with finite sets of input and output symbols, a finite set of states, a next-state function Ψ and an output function Φ:

$$\Psi: \quad (\underline{x}_j, \underline{s}_j) \rightarrow \underline{s}_{j+1} \ ,$$

$$\Phi: \quad (\underline{x}_j, \underline{s}_j) \rightarrow z_j \ ,$$

where $\underline{x}_j = [x_{1,j}, x_{2,j}, \ldots, x_{n,j}]$ and where $\underline{s}_j = [s_{1,j}, s_{2,j}, \ldots, s_{k,j}]$ and $\underline{s}_{j+1} = [s_{1,j+1}, s_{2,j+1}, \ldots, s_{k,j+1}]$ are the state vectors with the k q-ary components $s_{1,j}, s_{2,j}, \ldots, s_{k,j}$ and $s_{1,j+1}, s_{2,j+1}, \ldots, s_{k,j+1}$ at time instants j and j+1, respectively. $\underline{s}_0$ denotes the initial state. Fig. 1 shows a canonical representation for a FSM [1], driven by n q-ary sources.

H.C. Williams (Ed.): Advances in Cryptology - CRYPTO '85, LNCS 218, pp. 273-279, 1986.

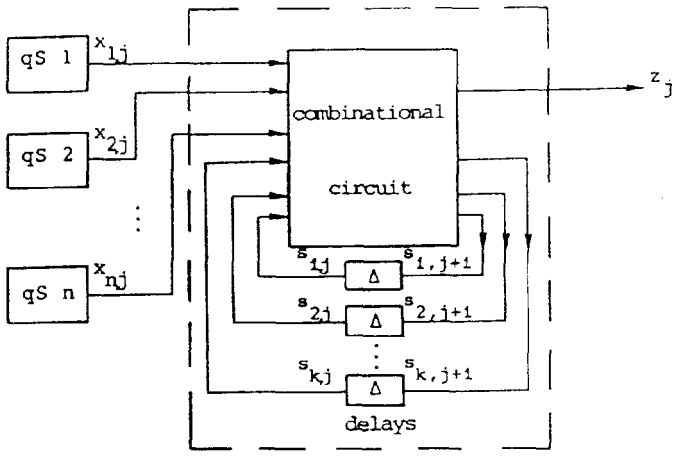


Fig. 1. A running key generator for use in stream ciphers.

A cryptanalyst possibly tries to break the above system by breaking the individual subkeys of the n sources. To prevent such divide and conquer attacks, the symbols generated by the FSM should be statistically independent on the symbols of one (or several) input sequences. In this note we give some results for FSM combiners.

II Correlation-Immunity of FSM combiners

A FSM combiner is called m-th order correlation-immune [2] if the mutual information between the running-key sequence z^j and every subset of m input sequences $x^j_{i1}, x^j_{i2}, \ldots, x^j_{im}$, $1 \leq i1 < i2 < \ldots < im \leq n$, is zero.

$$I(z^j ; x^j_{i1}, x^j_{i2}, \ldots, x^j_{im}) = 0 \quad , \text{ for all } j \geq 0, \tag{1}$$

where the superscript j means that all symbols up to time instant j are considered, e.g. $z^j = z_0, z_1, z_2, \ldots, z_j$. Note that z^j contains j+1 symbols. The sequences $x^j_{i1}, x^j_{i2}, \ldots, x^j_{im}$ are assumed to be independent of each other. Definitions (1) and (2) which is slightly stronger have been used by Rueppel [3,4].

$$I(z_j ; z^{j-1}, x^j_{i1}, x^j_{i2}, \ldots, x^j_{im}) = 0 \quad , \text{ for all } j > 0, \tag{2a}$$

and

$$I(z_j ; x^j_{i1}, x^j_{i2}, \ldots, x^j_{im}) = 0 \quad , \text{ for } j = 0 \; . \tag{2b}$$

In this section it will be shown that (2) is too restrictive to be used as a definition in general but is useful for a special (but cryptographically significant) case. Moreover, an expression equivalent to that given in (1) is derived. For ease in notation we assume m=1, however, the result is easily extended to any m, $1 \leq m < n$. From (2a) we obtain

$$I(z_j;z^{j-1},x_1^j) = I(z_j;z^{j-1}) + I(z_j;x_1^j|z^{j-1}) = 0 \ , \ j > 0. \tag{3}$$

Because mutual information is always positive, we must have

$$I(z_j;z^{j-1}) = H(z_j) - H(z_j|z^{j-1}) = 0 \ , \ j > 0, \tag{4}$$

and

$$I(z_j;x_1^j|z^{j-1}) = 0, \ j > 0. \tag{5}$$

For stationary input sequences (4) means that an independence definition according to (2) implies an independent and identically distributed (i.i.d.) sequence $z_0,z_1,\ldots$ which, of course, isn't necessary for correlation-immunity. Fig. 2 gives an example for the restriction made with a definition according to (2). All variables are binary and we assume in this example that the input sequences are balanced and i.i.d.

Example 1:

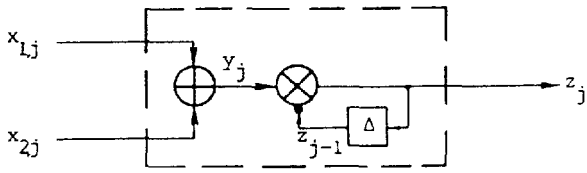


Fig. 2. A correlation-immune FSM with $I(z^j;x_i^j)=0$ but $I(z_j;z^{j-1},x_i^j)>0$ for i=1,2. '•' denotes inversion.

We certainly have $I(x_i^j;y^j)=0$ because the mod 2 addition at the input acts as a binary symmetric channel. From the data processing lemma follows that $I(x_i^j;z^j) \leq I(x_i^j;y^j) = 0$, for i=1,2. On the other hand, from $z_{j-1}=1$ follows that $z_j=0$, independently of the actual inputs. But this shows that $H(z_j|z^{j-1})<H(z_j)$ or $I(z_j;z^{j-1})>0$ from which follows that $I(z_j;z^{j-1},x_i^j)>0$ for i=1,2.

Now we proof the following equality:

$$I(z^j;x_1^j) = \sum_{i=1}^{j} I(z_i;x_1^i|z^{i-1}) \ , \ j > 0. \tag{6}$$

First we have

$$I(x_1^j;z^j) = I(z^{j-1};x_1^j) + I(z_j;x_1^j|z^{j-1}) \quad , j > 0. \tag{7}$$

The first term on the right hand side can be written as

$$\begin{aligned} I(z^{j-1};x_1^j) &= H(x_1^j) - H(x_1^j|z^{j-1}) \quad , j > 0. \\ &= H(x_1^j) - H(x_1^{j-1}|z^{j-1}) - H(x_{1j}|x_1^{j-1}z^{j-1}) \quad , j > 0. \end{aligned} \tag{8}$$

From the independence of the input sequences and the additional assumption that the initial state $\underline{s}_0$ is chosen independently of the input sequences, we have

$$\begin{aligned} H(x_{1,j}|x_1^{j-1}z^{j-1}) &= H(x_{1,j}|x_1^{j-1}\Phi'(x_1^{j-1},\ldots,x_n^{j-1},\underline{s}_0)) = \\ &= H(x_{1,j}|x_1^{j-1}) \quad , j > 0 \quad , \end{aligned}$$

and therefore it follows that

$$I(z^{j-1};x_1^j) = H(x_1^j) - H(x_1^{j-1}|z^{j-1}) - H(x_{1j}|x_1^{j-1}) \quad , j > 0.$$

The first term on the right hand side can be expanded as $H(x_{1j}|x_1^{j-1}) + H(x_1^{j-1})$ and therefore

$$I(z^{j-1};x_1^j) = H(x_1^{j-1}) - H(x_1^{j-1}|z^{j-1}) = I(z^{j-1};x_1^{j-1}) \quad , j > 0. \tag{9}$$

It follows from (7) and (9) that

$$I(x_1^j;z^j) = I(x_1^{j-1};z^{j-1}) + I(z_j;x_1^j|z^{j-1}) \quad , j > 0 \ . \tag{10}$$

(10) can be used iteratively to get (6). This completes the proof. From (6) immediately follows that the expressions given in (11) below are equivalent to expression (1) and therefore, are an equivalent definition for correlation-immunity of FSM's.

$$\begin{aligned} &I(z_j;x_{i1}^j,x_{i2}^j,\ldots,x_{im}^j|z^{j-1}) = 0 \text{ for all } j > 0 \\ \text{and} \quad &I(z_j;x_{i1}^j,x_{i2}^j,\ldots,x_{im}^j) = 0 \quad \text{for } j = 0 \ , \end{aligned} \tag{11}$$

where m and i1,i2,...,im in (11) are defined as in (1). Note also that the independence definitions (1) and (2) are equivalent if and only if the FSM generates an i.i.d. output sequence.

III A design criterion for finite state machines

In practice it may often be difficult to work with expression (1). In this section we assume that the input sequences are independent of each other and i.i.d. and we work out a much simpler condition.

Theorem:

A sufficient condition for (1) to hold is that the current state $\underline{s}_j$ and every set of m current inputs $x_{i1,j}, x_{i2,j}, \ldots, x_{im,j}$, $1 \leq i1 < i2 < \ldots < im \leq n$, are jointly statistically independent of the current output symbol z_j. If the FSM is a finite output memory machine which, moreover, generates an i.i.d. output sequence this condition is also necessary.

Note that it is sufficient due to the above theorem to fulfil some requirements on the memoryless output-function Φ independently on the chosen next-state function Ψ. To avoid unnecessary difficulties in notation the proof is given again for m=1 but is easily extended to any m, $0 < m < n$. First we have

$$H(z^j|x_1^j) = H(z_0|x_1^j) + H(z_1|z_0x_1^j) + \ldots + H(z_j|z^{j-1}x_1^j)$$

for a causal system with i.i.d. input sequences follows

$$H(z^j|x_1^j) = H(z_0|x_1^0) + H(z_1|z_0x_1^1) + \ldots + H(z_j|z^{j-1}x_1^j) \ .$$

For the FSM of Fig. 1 we have

$$H(z^j|x_1^j) \geq H(z_0|x_{1,0}) + H(z_1|\underline{s}_1x_{1,1}) + \ldots + H(z_j|\underline{s}_jx_{1,j})$$

or

$$H(z^j|x_1^j) \geq H(z_0|x_{1,0}) + \sum_{i=1}^{j} H(z_i|x_{1,i},\underline{s}_i) \ . \qquad (12)$$

Note that for a finite output memory machine (where the state is identical to some finite number of output digits) equality holds in (12). Now we use

$$I(z^j;x_1^j) = H(z^j) - H(z^j|x_1^j)$$

and together with (12) we obtain

$$I(z^j;x_1^j) \leq H(z^j) - H(z_0|x_{1,0}) - \sum_{i=1}^{j} H(z_i|x_{1,i},\underline{s}_i) \ .$$

The right hand side can be further increased by using

$$H(z^j) \leq \sum_{i=0}^{j} H(z_i) \ , \qquad (13)$$

and therefore

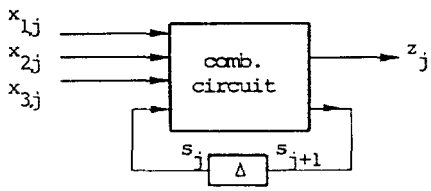
$$I(z^j;x_1^j) \leq H(z_0) - H(z_0|x_{1,0}) + \sum_{i=1}^{j} [H(z_i) - H(z_i|x_{1,i},\underline{s}_i)]$$

or

$$I(z^j;x_1^j) \leq I(z_0;x_{1,0}) + \sum_{i=1}^{j} I(z_i;x_{1,i},\underline{s}_i) \ , \qquad (14)$$

where equality holds in (14) for a finite output memory machine which fulfills (13) with equality. The theorem follows immediately from the fact that $I(z_i;x_{1i},\underline{s}_i) = 0$ is equivalent to saying that the current input $x_{1,i}$ and the current state $\underline{s}_i$ are jointly statistically independent of the current output z_i. Note that the FSM of Example 1 fulfills (1) even if the state and some inputs are not jointly statistically independent of the output. However, this is not a contradiction to the theorem because the finite output memory machine of Fig. 1 doesn't generate an i.i.d. sequence and therefore the necessary part of the theorem doesn't hold. The sequences in the following two examples have digits in GF(2).

Example 2:



$$s_{j+1} = \Psi(x_{1j},x_{2j},x_{3j},s_j)$$

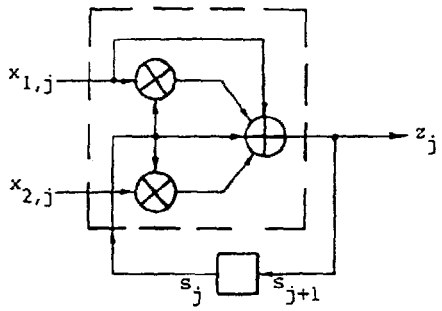
$$\Phi: z_j = x_{1j} \oplus x_{2j} \oplus s_j \cdot x_{2j} \oplus s_j \cdot x_{3j}$$

Fig. 3 A correlation-immune FSM with n=3, m=1.

The above FSM is correlation-immune with m=1 for any choice of Ψ due to the theorem of this section because x_{ij},s_j are jointly statistically independent of z_j for i = 1,2,3. (For every choice of $x_{i,j},s_j$ the output z_j is independently determined by the j-th digit of an i.i.d. sequence.)

Example 3:

The JK-FlipFlop (see Fig.4 for a logic equivalent) is an example for a finite output memory machine which generates an i.i.d. sequence when driven by two i.i.d. input sequences. However, it doesn't fulfil the necessary condition given in the theorem, as can be seen from the corresponding function Φ. For $s_j=0$ and any choice of $x_{1,j}$ we have $z_j=x_{1,j}$ and therefore s_j and $x_{1,j}$ are not jointly statistically independent of z_j.



$$\Phi: \quad z_j = x_{1j} \oplus s_j \oplus x_{1j} \cdot s_j \oplus x_{2j} \cdot s_j$$

$$\Psi: \quad s_{j+1} = z_j$$

Fig. 4. A finite output memory machine which generates an i.i.d. output sequence but is not correlation-immune.

Conclusions

Definitions for correlation-immunity of general finite state machines have been discussed. The input sequences have been assumed to be independent of each other. It turned out that the definitions according to (1) and (11) are equivalent. The definition according to (2) is equivalent to that given in (1) if and only if the output sequence generated by the FSM is an i.i.d. sequence. Further, a simple sufficient condition for FSM's to be correlation-immune has been developed under the assumption that the input sequences are independent of each other and i.i.d. . Moreover, it turned out that this condition is also necessary, if the FSM is a finite output memory machine which generates an i.i.d. output sequence.

Acknowledgment

The author is grateful to Dr. P. Schöbi from the Institute for Signal and Information Processing, Swiss Federal Institute of Technology, Zurich, for many helpful discussions and a thorough reading of the manuscript.

References

[1] R.E. Miller, "Switching Theory", Vol. II, Sequential Circuits and Machines, John Wiley & Sons, New York, London, Sydney, 1965.
[2] T. Siegenthaler, "Correlation-Immune Combining Functions for Cryptographic Applications", IEEE Tr. on Info. Theory,IT-30, No.5, Sept. 1984.
[3] R. Rueppel, "New Approaches to Stream Ciphers", Thesis, Swiss Federal Institute of Technology, No. 7714, 1984.
[4] ---, "How to Frustrate the Correlation Attack with one Bit of Memory" CRYPTO'85, Santa Barbara, Aug. 18-22, 1985.