STATISTICS SWEDEN
I/MET
Hans Block

REPORT

June 7, 1992

Cardinalities of the ranges

of iterated random functions

by

Hans Block, Statistics Sweden

Summary

Iterated random transformations on a finite set of n elements on itself are investigated. The question is how much the ranges will shrink. The expected cardinality of the ranges as well as the mean entropy of the transformations are estimated. The results are important for the construction of Message Authentication Codes, error propagation in block cryptos, transformations of cryptological keys, and random number generators.

Contents

1 Introduction and main result
2 Notations
3 Expected cardinalities
4 Entropies of the transformations
5 Some other iterated functions
6 File authentication
7 Error propagation in block cryptos
8 Transformation of keys
9 Random number generators
10 Conclusions
11 Acknowledgements

1 Introduction and main result

Cryptology is the science of very complicated functions on finite sets. These transformations are often - but not always - one-to-one. Even if the most important mapping in an algorithm is one-to-one, it can be built of bricks which are not injective. Message Authentication Codes (MAC) are cryptological transformations unto small sets, so they can not be injective. In the last sections we describe a few situations with non-injective cryptological functions.

If a transformation is not guaranteed to be injective, it almost certainly is not. In that case its range should be investigated. This is most important when non-injective functions are composed many times. The iterated function can degenerate seriously, so that a cryptanalyst can crack the code in some way.

In most cases, cryptological transformations are so complex or use such unusual elements that no explicit formulas can tell anything about their properties. Although the transformations certainly are deterministic, statistical models can and must be used.

We will study functions on finite sets which can be considered as truly random, i. e. the picture of each element is a stochastic variable with a uniform distribution, and all these stochastic variables are independent. If such types of functions are iterated many times, the range of the composed functions will be smaller and smaller. The rate of the shrinking is given by the following theorem:

Theorem. Given j transformations of a set of n elements into itself, where the pictures of all elements are uniformly distributed stochastic variables, and all such variables are independent. Then m_j, the expected value of the ratio between the cardinalites and the range of the product transformation satisfies

$$(1) \qquad \frac{1}{j+1} < m_j < \frac{2e}{j+1} \qquad \text{for } 0 \le j \le n.$$

The expected entropy H_j of the iterated transformation satisfies

$$(2) \qquad \log \frac{n}{j+1} < E(H_j) < \log \frac{2en}{j+1}$$

The first two inequalities were discovered independently of each other. The first one is due to Lennart Brynielsson, the second one to Hans Block. The last two inequalities were found by Lennart Brynielsson.

After the proof of this theorem, we will discuss a few more deterministic functions and the consequences of shrinking ranges.

2 Notations

f_j	A random function in step j
$f^{(j)}$	A function composed by j independent random functions
$\|R\|$	The cardinality of the set R
n	$\|\text{domain } f\|$
ξ_j	$\|\text{range } f^{(j)}\|$
m_j	$E(\xi_j) / n$
$p_k^{(j)}$	$P(k$ elements left after $f^{(j)})$
$p_{ik}^{(j)}$	$P(i$ elements left after $f^{(j+1)} \mid k$ elements left after $f^{(j)})$
$\eta_i^{(j)}$	$\|\{k; f^{(j)}(k) = i\}\|$
φ_i	$\eta_i^{(j-1)}$
ψ_i	$\eta_i^{(j)}$

3 Expected cardinalities

The bounds in the theorem are tight, differing only by a small constant factor. They need rather technical computations, using the convexity of several functions and up to four terms in the Taylor series.

First we note the following. When k elements are left in the range, it is irrelevant how many steps were needed to come there. Thus $p_{ik}^{(j+1)}$ is independent of j, and we can write

$$p_{ik}^{(j+1)} = p_{ik}$$

For the same reason we can define the stochastic variable $\xi|k$:

$$\xi|k = |\text{range } f^{(j+1)}| \text{ if } |\text{range } f^{(j)}| = k,$$

the number of elements left after $f^{(j+1)}$, if k elements are left after $f^{(j)}$. In order to calculate $E(\xi|k)$, we see that

$$\text{P(a certain element is missing after } f^{(j+1)} \mid |\text{range } f^{(j)}| = k) =$$

$$= (1 - 1/n)^k$$

$$\text{P(a certain element exists after } f^{(j+1)} \mid |\text{range } f^{(j)}| = k) =$$

$$= 1 - (1 - 1/n)^k$$

Now $\xi|k$ is the sum of n stochastic variables, each of which is $= 1$ if a certain element exists, and else $= 0$. Each of these variables has the expected value

$$1 - (1 - 1/n)^k,$$

as we see from the probabilities above. Thus

$$E(\xi|k) = n \,.\, (1 - (1 - 1/n)^k)$$

On the other hand,

$$E(\xi|k) = \sum_i i \,.\, p_{ik}$$

by definition. So

$$E(\xi_{j+1}) = \sum_i i \,.\, p_i^{(j+1)} =$$

$$= \sum_i i \sum_k p_{ik} \,.\, p_k^{(j)} =$$

$$= \sum_k p_k^{(j)} \,.\, \sum_i i \,.\, p_{ik} =$$

$$= \sum_k p_k^{(i)} \,.\, n \,.\, (1 - (1 - 1/n)^k)$$

We have used the fact that f_{j+1} is independent of the other fs and the expressions for $E(\xi \mid k)$. We now estimate m_{j+1}, the expected portion of elements left after the function $f^{(j+1)}$:

$$m_{j+1} = E(\xi_{j+1}) / n = \sum_k p_k^{(j)} \,.\, (1 - (1 - 1/n)^k) =$$

$$= 1 - \sum_k p_k^{(j)} . \exp(k \,.\, \ln(1 - 1/n)) = 1 - E(\exp(\xi_j \,.\, \ln(1-1/n))) \leq$$

$$\leq 1 - \exp(E(\xi_j \,.\, \ln(1 - 1/n))) = 1 - \exp(m_j \,.\, n \,.\, \ln(1 - 1/n)) \leq$$

$$\leq 1 - \exp(-m_j \,.\, (1 + 1/n))$$

where the first inequality is Jensen's inequality, and the second one follows from

$$\ln(1 - 1/n) \geq - 1/n - 1/n^2,$$

which holds for $n \geq 2$. With $c = 1 + 1/n$ we get

$$m_{j+1} \leq 1 - \exp(-m_j \,.\, c)$$

The simplest thing to do now would be to forget the factor c and the inequality, and put

$$m_{j+1} = 1 - \exp(-m_j),$$

and take three terms in the Taylor series of the exponential function:

$$m_{j+1} - m_j = 1 - 1 + m_j - m_j^2/2 - m_j = -m_j^2/2$$

and substitute the difference by a derivative and get

$$m(j)' = -m(j)^2$$

$$m' / m^2 = -1/2$$

$$D(-1/m) = -1/2$$

$$1/m = j/2 + C$$

$$m_j = 2/(j + C')$$

$$m_0 = 0 \Rightarrow C' = 1$$

This result would be sufficient for drawing the conclusions in the later part of the report. However, we want to prove inequalities and therefore we must make the following rather tedious calculations. In order to get rid of the factor c we need the following proposition:

Proposition. Let H be a function with $H(0) = 0$ and

$$H'(x) \geq 0, H''(X) \leq 0 \qquad \text{for } x \geq 0.$$

Let c be a constant > 1, $\{m_j\}$ and $\{q_j\}$ sequences such that

$$m_0 = 1, q_0 = 1$$

$$m_{j+1} \leq H(c \cdot m_j) \qquad \text{for } j \geq 0$$

$$q_{j+1} \geq H(q_j) \qquad \text{for } j \geq 0$$

Then

$$m_j \leq c^j \cdot q_j \qquad \text{for } j \geq 0.$$

Proof. Induction. It is true for 0. Assume it is true for j. Then

$$m_{j+1} \leq H(c \cdot m_j) \leq H(c^{j+1} \cdot q_j) \leq$$

$$\leq c^{j+1} \cdot H(q_j) \leq c^{j+1} \cdot q_{j+1},$$

due to the induction assumption and the facts that H is increasing, m and q satisfy the conditions above and that H satisfies

$$H(c \cdot x) \leq c \cdot H(x) \qquad \text{for all } c > 1,$$

which is a simple convexity property. ♠

We define

$$H(x) = 1 - exp(-x)$$

and note that it satisfies the assumptions of the proposition. We want to estimate the sequence $\{q_j\}$ defined by

$$q_0 = 1$$

$$q_{j+1} = H(q_j)$$

Numerical experiments or heuristic reasoning suggest

$$q_j < 2 / (j+1).$$

In order to prove that, we put

$$q_j = (2 - d_j)/ (j+1)$$

$$d_j = 2 - q_j \,.\, (j+1)$$

and get the new recurrence formula

$$d_0 = 1$$

$$d_{j+1} = 2 - (j+2) \,.\, (1 - exp((d_j - 2) / (j+1)))$$

We want to prove by induction that

$$0 \le d_j \le 1 \qquad for\, j \ge 0$$

Proof. It is true for 0. Assume it is true for j. Then

$$d_{j+1} \le 2 - (j+2) \,.\, (1 - exp(-1/(j+1))) \le$$

$$\le 2 - (j+2) \,.\, (1/(j+1) - (1/(2 \,.\, (j+1)^2)) =$$

$$= 1 - (j / (2 \,.\, (j+1)^2) \le 1$$

and

$$d_{j+1} \geq 2 - (j+2) \,.\, (1 - \exp(-2/(j+1))) \geq$$

$$\geq 2 - (j+2) \,.\, (2/(j+1) - 4/(2 \,.\, (j+1)^2) + 8/(6 \,.\, (j+1)^3)) =$$

$$= 2 \,.\, (j-1)/(3 \,.\, (j+1)^3) \geq 0 \qquad \text{for } j > 0.$$

Here we have used the induction assumptions and that the exponential function is increasing and satisfies

$$x - x^2/2 < 1 - \exp(-x) < x - x^2/2 + x^3/6$$

for $0 \leq x \leq 4$. Since it is easily checked that $d_1 > 0$, we have shown that

$$0 \leq d_j \leq 1 \qquad \text{for } j \geq 0$$

(In fact we could have shown, with the same type of calculations, that

$$d_j \approx (2 \,.\, \ln j) / 3j \qquad \text{as } j \to \infty$$

but that is not interesting for us.) ♠

Thus, using the proposition and the estimate for q_j we have proved

$$m_j < (1 + 1/n)^j \,.\, 2 / (j+1) < 2e / (j+1) \quad \text{for } 0 \leq j \leq n$$

which gives the second inequality in *(1)*.

4 Entropies of the transformations

In order to prove the other inequality in *(1)*, we study the stochastic variables $\eta_i^{(j)}$ which are defined to be the number of elements mapped onto i after j steps:

$$\eta_i^{(j)} = |\{k; f^{(j)}(k) = i\}|$$

We first calculate their expectation and variance and we will use recurrence relations. In order to simplify the notations we put

$$\varphi_i = \eta_i^{(j-1)} \text{ and } \psi_i = \eta_i^{(j)}$$

From symmetry it follows that all φ_i have the same distribution and all ψ_i another. Since

$$\sum_{i=0}^{n-1} \varphi_i = \sum_{i=0}^{n-1} \psi_i = n$$

they are dependent. The definitions of φ_i, ψ_i, and $f^{(j)}$ give the recurrence formula

$$\psi_i = \sum_{k=0}^{n-1} \chi_k \cdot \varphi_k$$

where the χ_k are stochastic *0-1* variables, independent with each other and with the φ_k, satisfying

$$P(\chi_k = 1) = 1/n.$$

Now we have

$$E(\psi) = E(\psi_i) = \sum_{k=0}^{n-1} E(\chi_k) \,.\, E(\varphi_k) = n \,.(1/n)\,.\, E(\varphi) = E(\varphi)$$

so that the expectation is constant and equals *1*, since $\eta_k^{(0)} \equiv 1$. For the second moment we get

$$E(\psi^2) = E(\sum_k \chi_k^2 \varphi_k^2) + E(\sum_{j \neq k} \chi_j \cdot \chi_k \cdot \varphi_j \cdot \varphi_k) =$$

$$= \Sigma\, (1/n) \cdot E(\varphi^2) + \sum_{j \neq k} (1/n^2) \cdot E(\varphi_j \cdot \varphi_k) =$$

$$= n \cdot (1/n - 1/n^2)\, E(\varphi^2) + (1/n^2) \cdot E(\Sigma\, \varphi_k)^2 =$$

$$= 1 + (1 - 1/n)\, E(\varphi^2)$$

which gives

$$Var(\psi) = (1 - 1/n) \cdot E(\varphi^2) = (1 - 1/n)(1 + Var(\varphi))$$

so the variance of step j is

$$\sum_{k=1}^{j} (1 - 1/n)^k < j$$

Next we consider the distribution

$$\eta_i^{(j)} / n, \qquad i = 0, 1, \ldots, n\text{-}1.$$

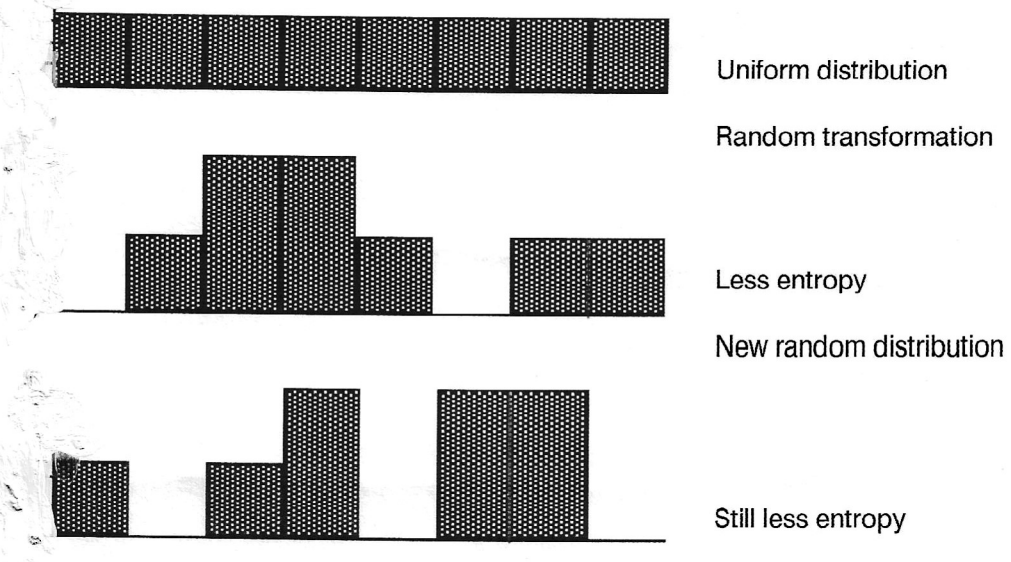

Figure 1. The number of elements mapped on i after 0, 1 and 2 random transformations. Since $\Sigma\, \eta_i = n$, the numbers $\eta_i^{(j)} / n$, $i = 0, 1, \ldots, n\text{-}1$ will give a probability distribution. We can compute ist entropy, which will decrease with each iteration. Since there are ξ_j elements > 0, the entropy must be $< log\ \xi_j$.

The expected entropy is

$$E(H_j) = E\Big(\sum_{i=0}^{n-1} -\frac{\eta_i^{(j)}}{n} \log \frac{\eta_i^{(j)}}{n}\Big) = \log n - E(\eta^{(j)} \log \eta^{(j)})$$

The term

$$E(\eta \log \eta) = \sum_{k=0}^{n-1} p_k k \log k,$$

and since $E(\eta) = \Sigma\, p_k k = 1$, we can use the concavity of the logarithm and conclude

$$E(\eta \log \eta) = \Sigma\, p_k k \log k \le \log \Sigma\, p_k k^2 =$$

$$= \log E(\eta^2) < \log(j+1)$$

Thus we have established the inequality

$$E(H_j) > \log \frac{n}{j+1}$$

Now since the distribution $\eta_i^{(j)}/n$ contains ξ_j non-zero elements, we know that its entropy must be less than $\log \xi_j$. This together with the fact that the logarithm function is increasing and convex gives

$$E(H_j) < E(\log \xi_j) < \log E(\xi_j)$$

and

$$\frac{n}{j+1} < E(\xi_j)$$

Combining this with the result from the previous section, we have an upper and a lower bound for m_j:

$$\frac{1}{j+1} < m_j < \frac{2e}{j+1} \qquad \text{for } 0 \le j \le n.$$

and also for the expected entropy $E(H_j)$:

$$\log \frac{n}{j+1} < E(H_j) < \log \frac{2en}{j+1}$$

so we have concluded the proof of our main theorem. ♠

5 Some other iterated functions

The assumption in the theorem above are rather strong, but we have got similar results for much more deterministic functions $f^{(i)}$. An example is

$$f(i) = i^2 + a \quad (\text{mod } p), \quad p \text{ prime.}$$

Although the range shrinks to 1/2 in the first iteration, the function behaves pretty well up to a point, after which the range is constant.

On the other hand, if the transformation is systematic in some way, the decrease in cardinality could be much worse. The same operation modulo a composite number is one example of this.

A much worse example is given by nilpotent linear transformations on finite fields. They reduce the *dimension* of the range with one unit in each step, so after r iterations, (where r is the dimension of the domain), the range is the single vector 0. Another very bad one is given by the following function:

Let f consist of two steps. The first step is a projection which maps 1/3 of all elements on one single element, and all other elements on themselves. The second step is a random permutation. Compose many such double transformations.

In the first step, the range is reduced by a factor 2/3 which is greater than the factor $(1 - 1/e)$ for truly random transformations. But in this case, the second, third, ... iterations will also reduce the range with almost a factor 2/3. Thus, the range will be reduced to a single point within $O(\ln n)$ iterations instead of $O(n)$ iterations in the truly random case.

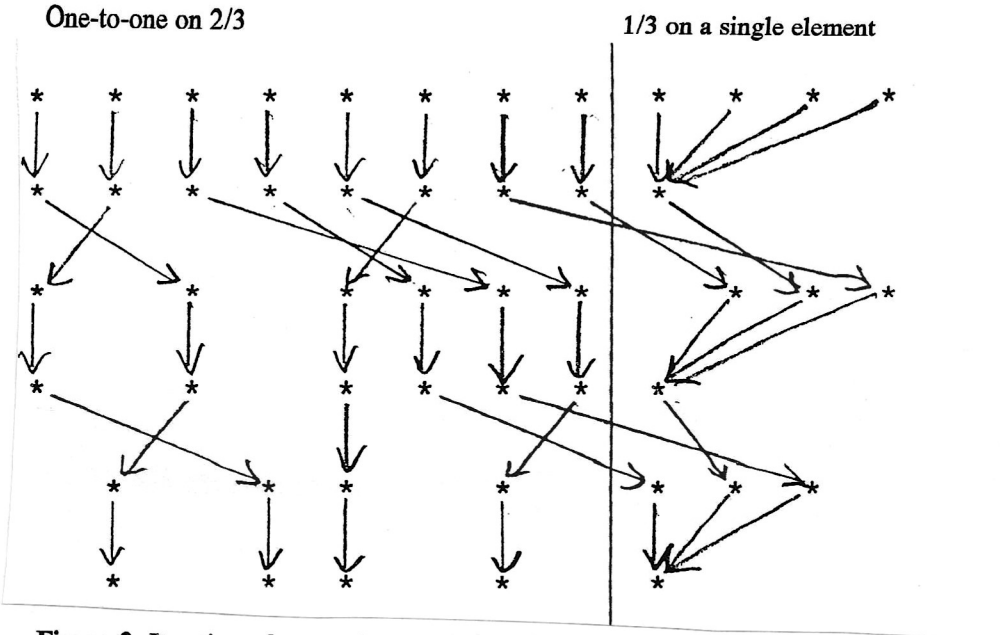

Figure 2. Iteration of a mapping consisting of a random permutation and a projection of 1/3 of the elements on one element. The reduction of the ranges goes much faster since many elements are mapped on the same element by each f_j.

In fact,

$$E(\xi \mid k) = 2k / 3 + 1 - (2/3)^k$$

for this transformation, and it is easy to show that

$$E(\xi_{j+1}) < (2/3) \cdot E(\xi_j) + 1$$

so that

$$E(\xi_j) - 3$$

decreases at least geometrically.

6 File authentication

6.1 Definitions

An authenticating algorithm is a complicated function F from a data file T, also called the plaintext, and a secret key K to a number S, called the Message Authentication Code (MAC):

$$S = F(T, K)$$

The file T is a bitstream of arbitrary length. K and S are non-negative integers less than certain constants. F is usually not one-to-one in either argument. The file may consist of a single transaction. However, in this paper we are interested of MACs on long files.

MACs are used to protect files during transportation or storage. The key is secret and known only to the sender and receiver. The sender computes the MAC and appends it to the file.

The receiver also computes the MAC and compares the newly computed MAC with the MAC of the file. If they agree, he can be sure that:

- The file comes from the proper sender
- The file has not been changed during the transportation or storage.

6.2 Security claims

In order to fulfill these ends, a good algorithm F should have the following properties:

- The algorithm should be safe, provided the key is secret. For cryptanalysis, the algorithm must be regarded as known.

The algorithm should be safe in the following respects:

- There should be no way to change an intelligible plaintext without changing the MAC.
- There should be no possibility to get the key from a number of given plaintext files and their corresponding MACs.

The authentication program module with a loaded key should always be protected against unauthorized use. To be sure against that threat too, we add a third claim:

- It should be very difficult to get the key from a number of plaintexts chosen by the cryptanalyst and their corresponding MACs.

We want to make these claims a little more concrete and demand the following properties of the algorithm:

- Both the key and the MAC should be sufficiently long.
- In the MAC, all digits should occur with the same probability.
- A change of one single bit in the plaintext should give a big change in the MAC.
- A change of one single bit in the key should give a big change in the MAC.
- A change of a single character in the plaintext should not be possible to compensate by a change of a few other characters in the plaintext.
- The MAC should be sensitive to insertions and deletions of single plaintext characters, even zeros and blanks.
- The MAC should be sensitive to permutations of the plaintext, both between adjoining characters and between long characters strings.
- All characters of the plaintext should contribute to the MAC in equal degree, i. e., the algorithm should be symmetric in the plaintext.
- The authentication transformation should be locally injective in the following sense: If two plaintext files are equal but for a small segment anywhere in the files, there should be a high probability that the corresponding MACs are different.
- The above claims should hold even for weak keys, consisting of for example only zeros.
- The transformations should be non-linear.
- No step of the algorithm should use only a small part of the key.
- The algorithm should be so simple that both a cryptological review and an efficient program checking are possible.

6.3 Constructing an algorithm

File authentication algorithms can be constructed in many ways. Limitations of storage and computing time put certain restrictions. Most algorithms are made in the following way:

Let the plaintext file T be divided into blocks

$$t_j, \qquad 0 \le t_j < n, j = 1, \ldots, J$$

of equal length, where the last block has to be padded in some predefined way, if necessary. The function F is determined as follows:

$$s_0 = g(K)$$

$$s_j = f(s_{j-1}, t_j, K), \qquad j = 1, \ldots, J$$

$$S = F(T) = h(s_J, K)$$

f, g and h are some functions, and the s_js, which we store in an *accumulator*, are non-negative integers less than n. The computing time depends mainly on F.

If f is complicated enough, most of the claims above are fulfilled. The introduction of new plaintext adds a stochastic element to the function of the first argument. The claim on local injectivity tells us that f should be injective in the second argument.

If f is not injective in the first argument, the following could happen:

We keep the key and all plaintext blocks but the first one fixed. We vary the first plaintext block in all possible ways. We consider the corresponding values of the accumulator s_1, and transform this set successively by the functions

$$f_1, f_2, \ldots, f_J.$$

We consider the cardinalities of the ranges of the successive compositions of the f_js. The cardinalities form a nonincreasing sequence of integers, which could go down to very low numbers. If so, the first plaintext block could be changed in many ways, still giving the same MAC. The same is true if a great part of the beginning of the file is changed, provided that the rest of the file is sufficiently long. The algorithm tends to "forget" the beginning of the file. It violates several of the claims above, making the MAC useless.

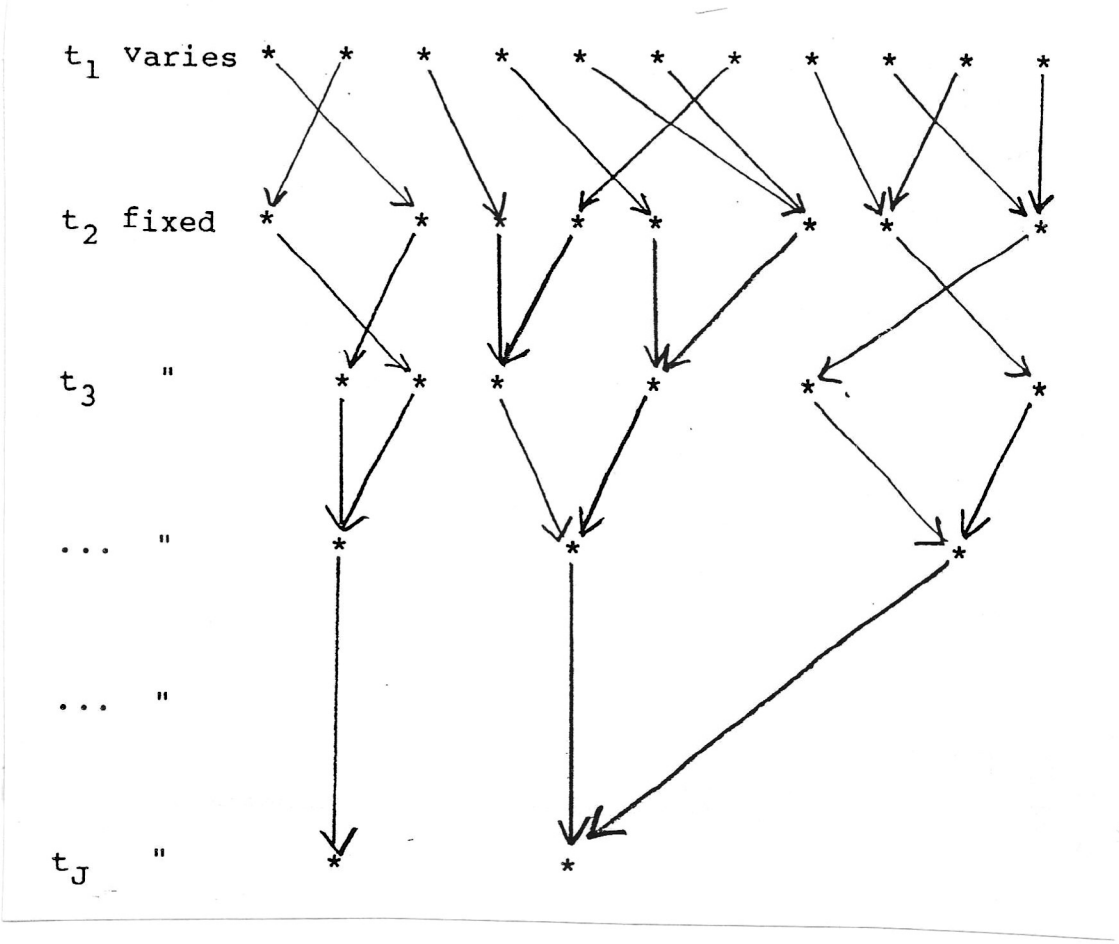

Figure 3. The risk with shrinking ranges for MACs. The number of possible pictures of the fist block may be too small after many iterations of the function f. If so, it will be possible to change the first part of the file *T*, and still have a high probabilty that the MAC is unchanged.

Our investigation tells that this danger is not serious, if truly random functions are used so the model is applicable, and the accumulator is big enough. On the other hand, in banking applications 1 000 000 blocks would be quite realistic. If $n = 2^{64}$, the number of bits in the accumulator is reduced by almost 1/3. It seems impractical to do lots of computations on big numbers, when only much smaller numbers are irrelevant.

7 Error propagation in block cryptos

Another application of our results is error propagation in blockcryptos. Let the plaintext consist of blocks

$$t_j, \qquad 0 \le t_j < n, j = 1, \ldots, J$$

and let the encrypted blocks be

$$s_j = f(s_{j-1}, t_j, K), \qquad j = 1, \ldots, J$$

where

$$s_0 = g(K).$$

As in the authentication case, an error in a block in the beginning of the plaintext might not propagate to the end of the cryptotext, if f is random in the first argument, n rather small and the plaintext is long.

This could be bad in some cases.

8 Transformations of keys

A cryptographic algorithm may have many steps, each of which needs many digits from a key. If the key should not be too long, in order not to annoy the user, the key has to be used many times. In order to prevent further analysis, when one step has been cracked, the key must be transformed between the steps. A non-injective transformation could prevent analysis between the steps.

The range of the transformation must be big enough, allowing sufficiently many choices for the keys. Our investigation shows that the range does not shrink too much, if the transformation can be regarded as random.

9 Random number generators

In many cases random numbers are generated by a formula

$$x_{j+1} = f(x_j)$$

where the x_js belong to a set of cardinality n. The important thing is to get a long period. What could be said about the period, if $f = f_j$ are chosen to be random functions?

If they are truly random, they will not repeat, and the random numbers will not repeat. If they are pseudo-random, the periodicity of the generator will be at least as long as the period of the functions, so our results have no relevance.

But if the purpose is only to generate random numbers, it would be much more difficult to generate many random functions, so the constructor would be tempted to use the same function in all steps, but choose it in some random-like way. Then the random numbers will have a period of length at most n. It is well-known, that the length will probably be only $O(\sqrt{n})$, according to the birthday problem. This is true even if the function f is one-to-one.

Our model is certainly not applicable in this case, but it still indicates that the loss of range according to non-injective functions is much less serious than the risk of meeting an element twice.

The conclusion is that the constructor must guarantee the length of the period in some way, but can choose a function which is not injective.

10 Conclusions

We have got very exact estimates of the expected cardinality of the range of iterated random functions. The ranges shrink with a factor

$$C/j$$

where j is the number of iterations.

In most cases this is not very dangerous. The exception is authentication of very large files using small accumulators. However, it is essential that the constructor of algorithms checks that the probabilistic model can be used.

11 Acknowledgements

The results on the entropy, including the first inequality for the expected cardinality, was given by Dr Lennart Brynielsson, Department of Communications Security, Swedish Defence Staff. He has also read the text and given many valuable comments. The late Professor Tore Herlestam has read a version of the paper and simplified a proof. The work was supported by SÄKdata AB, in the development of the Electronic Seal, a file authentication algorithm which is a banking standard in Sweden an Finland.

The material was presented at Eurocrypt 1984, but was never published.