RC 20674 (12/24/96)
Mathematics

IBM Research Report

Luby-Rackoff: Four rounds is not enough

D. Coppersmith

IBM Research Division
T.J. Watson Research Center
Yorktown Heights, New York

IBM Research Division
Almaden · T.J. Watson · Tokyo · Zurich

Luby-Rackoff: Four rounds is not enough

Don Coppersmith
IBM Research
T. J. Watson Research Center
Yorktown Heights, NY 10598, USA

1 Introduction

Luby and Rackoff [3] present a framework for analyzing generalized Feistel ciphers. These are ciphers on $2n$-bit messages, which proceed by breaking the message into n-bit halves and, for a certain number of *rounds*, iteratively replace one message half by its exclusive-or with some *round function* f_i of the other half. They show that with a three-round cipher, it requires $\Omega(2^{n/2})$ chosen plaintexts to distinguish the cipher from a random permutation on $2n$-bit messages; and with a four-round cipher, even if chosen plaintexts and chosen ciphertexts are allowed, it will require $\Omega(2^{n/2})$ chosen texts to distinguish the cipher from a random permutation.

A four-round Luby-Rackoff cipher on $2n$-bit messages has four "round functions" $f_1, \ldots, f_4$, containing 4×2^n (n-bit) words or $4n \times 2^n$ bits of information. Thus, from an information-theoretic standpoint, we need at least 2×2^n pairs of plaintext and corresponding ciphertext in order to recover the contents of the round functions. In the present paper we show a computationally efficient attack, with a chosen-plaintext requirement very close to this lower bound: with $O(2^n)$ chosen plaintexts we can recover most of the round functions f_i.

Our results are a companion to those of Aiello and Venkatesan [1], who use only $O(2^{n/2})$ chosen plaintexts to distinguish between a four-round Luby-Rackoff cipher and a random permutation on $2n$-bit messages.

The rest of this paper is organized as follows. In Section 2 we set forth the Luby-Rackoff cipher. In Section 3 we present Aiello and Venkatesan's attack, which will motivate the present work. Section 4 gives the details of the present attack. Section 5 concludes with some open problems. A discussion on our random graph model is deferred to the appendix, Section 6.

2 The Luby-Rackoff cipher

There are four "round functions" $f_i : \mathbf{Z}_2^n \to \mathbf{Z}_2^n$, each being a secret random (or pseudorandom) function or a secret (pseudo)random permutation.

Encryption proceeds as follows.

$$\begin{aligned} \text{Plaintext} &= M_0||M_1 \\ M_{i+1} &= M_{i-1} \oplus f_i(M_i), \qquad i = 1, 2, 3, 4 \\ \text{Ciphertext} &= M_5||M_4, \end{aligned}$$

where $M_i \in \mathbf{Z}_2^n$, and $M_0||M_1$ denotes concatenation. The reversal of ciphertext $(M_5||M_4)$ is intended to enhance the operational similarity between encryption and decryption.

Decryption proceeds as follows.

$$\begin{aligned} \text{Ciphertext} &= M_5||M_4 \\ M_{i-1} &= M_{i+1} \oplus f_i(M_i), \qquad i = 4, 3, 2, 1 \\ \text{Plaintext} &= M_0||M_1. \end{aligned}$$

Aiello and Venkatesan [1] give a method for distinguishing a 4-round Luby-Rackoff function from a random permutation R_{2n} on $\mathbf{Z}_2^{2n}$, using $O(2^{n/2})$ chosen plaintexts and corresponding ciphertexts. They depend on the birthday phenomenon to find a pair (M^i, M^j) such that $M_1^i = M_1^j$ and $M_0^i \oplus M_4^i = M_0^j \oplus M_4^j$.

In the present note we give a new attack. In contrast to [1] we look for pairs (M^i, M^j) such that $M_1^i = M_1^j$ and $M_4^i = M_4^j$. We give a randomized algorithm to discover the round functions f_i (up to equivalence classes), not just to distinguish Luby-Rackoff from a random permutation. We use $O(n2^n)$ chosen plaintexts and corresponding ciphertexts, and $O(n2^n)$ computational steps and $O(2^n)$ space, to recover all of f_i with high probability. An alternate attack uses $O(2^n)$ chosen plaintexts, computational steps and space to recover 99.9% of the contents of f_i, enabling decryption of 99.6% of possible ciphertext.

2.1 Notation and Conventions

We will not distinguish between random and pseudorandom functions.

$\mathbf{Z}_2$ denotes the field with two elements.

M_i^j denotes an n-bit quantity used as the input on round i when encrypting the j-th message, for $i = 1, 2, 3, 4$. By extrapolation, M_0^j is part of the plaintext and M_5^j is part of the ciphertext.

2.2 Symmetries

Several different instances of the round functions (f_1, f_2, f_3, f_4) will lead to the same input-output behavior of the Luby-Rackoff cipher. That is, given a cipher $LR4$ defined by (f_1, f_2, f_3, f_4), and two arbitrary n-bit constants c and d, we can define an alternate cipher $LR4'$ by its round functions

$$\begin{aligned} f_1'(x) &= f_1(\quad x \quad) \oplus c \\ f_2'(x) &= f_2(x \oplus c) \oplus d \\ f_3'(x) &= f_3(x \oplus d) \oplus c \\ f_4'(x) &= f_4(\quad x \quad) \oplus d \end{aligned}$$

It is straightforward to check that $LR4$ and $LR4'$ will both map a given plaintext $(M_0||M_1)$ to the same ciphertext.

We will use these two degrees of freedom (c and d) to specify:

$$\begin{aligned} f_1(0) &= 0 \\ f_2(0) &= 0. \end{aligned}$$

This will simplify the description of our attack.

3 Aiello and Venkatesan's attack

Aiello and Venkatesan [1] present a method to distinguish a Luby-Rackoff cipher from a random permutation on $\mathbf{Z}_2^{2n}$, in the case where the round functions f_i are random functions. They require $O(2^{n/2})$ chosen plaintexts. We present their attack for background interest.

Present $O(2^{n/2})$ plaintext messages M^i with the same value of $M_1^i = M_1^*$, namely $(M_0^i||M_1^*)$, and obtain the corresponding ciphertexts. Look for instances where

$$M_0^i \oplus M_4^i = M_0^j \oplus M_4^j.$$

Since

$$M_0^i \oplus M_4^i = f_1(M_1^i) \oplus f_3(M_3^i)$$

and

$$M_1^i = M_1^j,$$

this implies

$$f_3(M_3^i) = f_3(M_3^j).$$

If the f_i are random functions, this will happen with higher probability than with a random function R_{2n} on $\mathbf{Z}_2^{2n}$. Namely, it will happen in either of the following two situations:

(1) $M_2^i \neq M_2^j$ but $f_2(M_2^i) = f_2(M_2^j)$ (so $M_3^i = M_3^j$); or

(2) $M_3^i \neq M_3^j$ but $f_3(M_3^i) = f_3(M_3^j)$.

If the f_i are random functions, the expected number of matches after $q = c2^{n/2}$ plaintexts is about

$$\binom{q}{2}\frac{2}{2^n} \approx \frac{q^2}{2^n} = c^2.$$

If the f_i are random permutations, neither condition (1) nor (2) can hold, so we will have no matches. Finally, the corresponding expected number for R_{2n} is

$$\binom{q}{2}\frac{1}{2^n} \approx \frac{q^2}{2 \times 2^n} = \frac{c^2}{2}.$$

This experiment serves to probabilistically distinguish 4-round Luby-Rackoff from a random permutation. The higher the value of c (and hence q), the better the discrimination.

4 Our attack

Rather than just distinguish a Luby-Rackoff cipher from a random permutation, we will attempt to discover the secret information in the cipher, namely the round functions f_i.

We also supply plaintext pairs with $M_1^i = M_1^j$, but we look for instances where

$$M_4^i = M_4^j.$$

From

$$\begin{aligned} M_1^i \oplus M_5^i &= f_2(M_2^i) \oplus f_4(M_4^i) \\ M_1^i &= M_1^j \\ M_4^i &= M_4^j, \end{aligned}$$

we deduce

$$\begin{aligned} M_5^i \oplus M_5^j &= f_2(M_2^i) \oplus f_2(M_2^j) \\ &= f_2\left[M_0^i \oplus f_1(M_1^i)\right] \oplus f_2\left[M_0^j \oplus f_1(M_1^j)\right]. \end{aligned}$$

Treating $f_1(M_1^i) = f_1(M_1^j)$ as an unknown constant b, we obtain

$$M_5^i \oplus M_5^j = f_2(M_0^i \oplus b) \oplus f_2(M_0^j \oplus b). \tag{1}$$

Remark: In order for this coincidence to happen, we must have

$$M_2^i \oplus f_3\left[M_1 \oplus f_2(M_2^i)\right] = M_2^j \oplus f_3\left[M_1 \oplus f_2(M_2^j)\right].$$

If $M_1^i = M_1^j$, M_2^i, and M_2^j are all randomly selected, and f_i are either random permutations or random functions, each such event will happen with probability about $1/2^n$.

We begin the attack. Fix a value of M_1, say $M_1^{1,*} = 0$. (The first superscript 1 will identify a family of plaintexts; the second superscript * reminds us that M_1 is shared among this family.) Define $b_1 = 0$. In general we will have $b_k = f_1(M_1^{k,*})$, but by our choices so far we have

$$b_1 = f_1(M_1^{1,*}) = f_1(0) = 0.$$

Present as chosen plaintext all 2^n texts of the form $(M_0^{1,i}||M_1^{1,*})$. Sort the output ciphertexts according to the value $M_4^{1,i}$. Assuming the round functions $f_1, \ldots, f_4$ are either random functions or random permutations, we expect to find, for each $k = 0, 1, 2, \ldots, n/\log_2 n$, about

$$\frac{e^{-1}}{k!}2^n$$

values m_4, each of which matches exactly k values of $M_4^{1,i}$. If $k \geq 1$ each such m_4 yields $k-1$ nonredundant equations of the form (1). Thus we come up with about

$$\sum_{k \geq 1} \frac{e^{-1}}{k!}2^n(k-1) = \left(\frac{1}{e}\right) 2^n$$

nonredundant equations

$$f_2(M_0^{1,i} \oplus b_1) \oplus f_2(M_0^{1,j} \oplus b_1) = M_5^{1,i} \oplus M_5^{1,j},$$

or equivalently, using $b_1 = 0$,

$$f_2(M_0^{1,i}) \oplus f_2(M_0^{1,j}) = M_5^{1,i} \oplus M_5^{1,j}. \tag{2}$$

Select one of the pairs (i_1, j_1) for which we have obtained equation (2). Define

$$\Delta = M_5^{1,i_1} \oplus M_5^{1,j_1},$$

and remark that we also have $\Delta = M_3^{1,i_1} \oplus M_3^{1,j_1}$. Define a new value of M_1, say $M_1^{2,*}$, by

$$M_1^{2,*} = M_1^{1,*} \oplus \Delta.$$

Again present all 2^n chosen plaintexts of the form $(M_0^{2,i}||M_1^{2,*})$, collect the corresponding ciphertexts, and sort according to the values of $M_4^{2,i}$, forming pairs as before.

There is one pair of particular interest. Consider the unique message $(M_0^{2,i_2}||M_1^{2,*})$ which satisfies

$$M_2^{2,i_2} = M_2^{1,i_1},$$

and the unique message with

$$M_2^{2,j_2} = M_2^{1,j_1}.$$

We do not know the values of i_2 and j_2 yet. We can deduce that

$$\begin{aligned} M_3^{2,i_2} &= M_1^{2,*} \oplus f_2(M_2^{2,i_2}) = M_1^{2,*} \oplus f_2(M_2^{1,i_1}) = \Delta \oplus M_1^{1,*} \oplus f_2(M_2^{1,i_1}) \\ &= \Delta \oplus M_3^{1,i_1} = M_3^{1,j_1} \end{aligned}$$

and similarly

$$M_3^{2,j_2} = M_3^{1,i_1}.$$

Then it is straightforward to compute

$$\begin{aligned} M_4^{2,i_2} &= M_4^{2,j_2} \\ M_5^{2,i_2} &= M_5^{2,j_2} \oplus \Delta \\ M_0^{2,i_2} &= M_0^{2,j_2} \oplus (M_0^{1,i_1} \oplus M_0^{1,j_1}). \end{aligned}$$

We expect to find only one pair (i_2, j_2) matching all these conditions, although we cannot distinguish the correct order; we cannot tell (i_2, j_2) from (j_2, i_2).

Putting this all together, and recalling $b_1 = 0$, we find two possible values for the quantity b_2:

$$b_2 = b_2 \oplus b_1 = \begin{cases} M_0^{1,i_1} \oplus M_0^{2,i_2} \\ M_0^{1,i_1} \oplus M_0^{2,j_2} \end{cases}.$$

For each of the possible values of b_2, we try to piece together the partial information on f_2 given by the equations

$$\begin{aligned} f_2(M_0^{1,i}) \oplus f_2(M_0^{1,j}) &= M_5^{1,i} \oplus M_5^{1,j} \\ f_2(M_0^{2,k} \oplus b_2) \oplus f_2(M_0^{1,\ell} \oplus b_2) &= M_5^{2,k} \oplus M_5^{2,\ell}. \end{aligned}$$

We will succeed in piecing them together if and only if we select the right order (i_2, j_2).

Continue the process, selecting a total of ℓ values $M_1^{k,*}$ and obtaining the encryption of $\ell 2^n$ chosen plaintexts, where we either choose $\ell = 5$ (for nearly complete coverage) or $\ell = n$ (for complete coverage).

Consider a graph on 2^n nodes, with an edge between x and y if, for some values i, j and for some k, $1 \leq k \leq \ell$, we have

$$\begin{aligned} x &= M_0^{k,i} \oplus b_k \\ y &= M_0^{k,j} \oplus b_k \\ f_2(M_0^{k,i} \oplus b_k) \oplus f_2(M_0^{k,j} \oplus b_k) &= M_5^{k,i} \oplus M_5^{k,j} \\ f_2(x) \oplus f_2(y) &= M_5^{k,i} \oplus M_5^{k,j} \end{aligned}$$

as one of the "nonredundant" equations above.

We would like to view this as a "random graph" (in the sense of Erdös and Rényi [2]) with $(\ell/e)2^n$ edges. In their model, if $\alpha > 1/2$ and N is large, if a graph G on N vertices has αN edges selected randomly from the $\binom{N}{2}$ possible positions, with high probability G has a "giant connected component" containing $N - o(N)$ vertices.

There are differences; considering only the $(1/e)2^n$ edges corresponding to one value of $M_1^{k,*}$, our process is biased against the formation of relatively large $O(n)$ connected components, as compared to the random graph. So the analysis of [2] cannot be applied directly; nonetheless we are confident that a similar analysis could be performed with similar results. Calculations based on our present model indicate that in our case the giant component is somewhat larger than that of the random graph case. See Section 6.

The results from random graphs (if they were applicable) would predict that our graph would have a giant connected component with $\gamma 2^n$ nodes where γ is a solution of

$$\begin{aligned} \gamma &= 1 - e^{-\alpha\gamma} \\ \alpha &= 2\ell/e \end{aligned} .$$

If $\ell = 5$ this component would contain about 0.972×2^n nodes, in the random graph model. Calculations indicate that with our model the percentage of coverage increases to 0.9925598.

ℓ	Random Graph	Our Model
2	0.56378	0∗
3	0.84512	0.90841
4	0.93646	0.97676
5	0.97201	0.99256
6	0.98720	0.99742
7	0.99402	0.99907

(In our model there is no giant component for $\ell = 2$.)

If $\ell = n$ this component would contain all 2^n nodes with high probability. To simplify discussion we assume that this component contains the node $x = 0$ for which we know the value of f_2: $f_2(0) = 0$. For each y in the connected component we can then obtain the value $f_2(y)$ by following the edges of the graph. Thus we obtain most or all of the values of the table f_2.

Having done this, the rest of the analysis is easy. Using our knowledge of f_2 and $f_1(M_1^{k,*}) = b_k$, we run through the existing data again to find f_3:

$$\begin{aligned} M_0^{k,i} \oplus M_4^{k,i} &= f_1(M_1^{k,*}) \oplus f_3(M_3^{k,i}) \\ &= b_k \oplus f_3\left[M_1^{k,*} \oplus f_2(M_0^{k,i} \oplus f_1(M_1^{k,*}))\right] \\ &= b_k \oplus f_3\left[M_1^{k,*} \oplus f_2(M_0^{k,i} \oplus b_k)\right] \end{aligned}$$

and f_4 similarly.

To find f_1 we need to submit another 2^n plaintexts, one for each value M_1. Partially decrypt the resulting ciphertext, using our known values of f_4, f_3 and f_2, to find the value of $f_1(M_1)$.

5 Conclusion

We have shown a chosen-plaintext attack against 4-round Luby-Rackoff ciphers. The quality of the results depends on the amount of chosen plaintext: with $n2^n$ samples, we can usually recover all of f_i (up to symmetry), while with 8×2^n samples we can already recover 99.9% of the entries of f_i. (This figure represents $\ell = 7$ families of 2^n plaintexts, plus the 2^n additional texts required for finding f_1.) This compares favorably to the information-theoretic lower bound of 2×2^n texts. The computation time and space are also bounded by $O(n2^n)$ (resp. $O(2^n)$).

An obvious open question is whether similar results could be achieved for a five-round cipher.

Another open question concerns the recursive Luby-Rackoff ciphers: if the round functions f_i are themselves Luby-Rackoff ciphers on n-bit messages (rather than $2n$-bit), so that they contain only $O(2^{n/2})$ words ($O(n2^{n/2})$ bits) of information, can we mount a similar attack using only $O(2^{n/2})$ chosen plaintexts?

6 Appendix: Random graph computations

In this section we sketch our computation of the size of the giant component.

Our model is that the graph on $N = 2^n$ vertices is composed of (the edges of) ℓ independent subgraphs, each of which has approximately $Ne^{-1}/h!$ components of size h, $h = 0, 1, \ldots, \log N$. In each subgraph, for a given random node x, the size of the component containing x (excluding x itself) is Poisson distributed with $\lambda = 1$. Throughout, we consider the asymptotics as $N \to \infty$. The probability that the component has size h (excluding x) is

$$\binom{N-1}{h} \left(\frac{1}{N}\right)^h \left(\frac{N-1}{N}\right)^{N-1-h} = \left(\frac{e^{-1}}{h!}\right) \left(1 + O(\frac{h^2}{N})\right).$$

We ask the probability that a random node x_0 in the whole graph is contained in a small component as opposed to the giant component. Let V_1 be the set of "neighbors" of x_0, nodes which lie in the same component as x_0 in one of the subgraphs, and inductively let V_{k+1} be the set of neighbors of the nodes in V_k, other than those already included in $\{x_0\} \cup V_1 \cup \ldots \cup V_k$. Set $v_k = |V_k|$.

We see that v_1 is Poisson distributed with $\lambda = \ell$; and if $v_1, \cdots, v_k$ are already known, then v_{k+1} is Poisson distributed with $\lambda = (\ell - 1)v_k$, because for each node in V_k, the neighbors associated with one subgraph have already been accounted for in $V_{k-1} \cup V_k$.

We selected a cutoff value $v_{\max}$ and created a Markov chain, with transition probabilities

$$m_{ij} = Prob\,(v_{k+1} = i | v_k = j) = \frac{[(\ell-1)j]^i}{i!} e^{-(\ell-1)j},$$

with absorbing barriers at $v_{\max}$ and 0. The initial vector was based on the Poisson distribution with $\lambda = \ell$. The final, steady-state solution was concentrated on the barriers, $i = 0$ and $i = v_{\max}$; we considered the latter condition to indicate that the node was in the giant component.

We found that for $\ell = 3, 4, \ldots, 7$, the computed probability of being in the giant component was independent of $v_{\max}$ in the range $13 \leq v_{\max} \leq 200$, to ten decimal places, so we accepted this as the actual probability. Already for $\ell = 7$ the complement of the giant component (0.000928) is dominated by isolated vertices (0.000912) and those in components of size two (0.000016).

For $\ell = 2$, we see that $E(v_{i+1}) = E(v_i) = E(v_1) = 2$, so there is no giant component, in contrast to the random graph case.

References

[1] W. Aiello and R. Venkatesan, "Benes: a non-reversible alternative to Feistel," *Advances in Cryptology - EUROCRYPT '96,* Ueli Maurer (Ed.) Springer LNCS **1070** (1996), 307–320.

[2] P. Erdös and A. Rényi, "On the evolution of random graphs," Magyar Tud. Akad. Mat. Kut. Int. Közl. 5 (1960) 17-61. Reprinted in P. Erdös, *The Art of Counting* (1973) pp. 574-618, and in *Selected Papers of Alfréd Rényi* (1976) pp. 482-525.

[3] M. Luby and C. Rackoff, "How to Construct Pseudorandom Permutations from Pseudorandom Functions," *SIAM Journal on Computing*, **17**, 373-386 (1988).

[4] M. Naor and O. Reingold, "On the construction of pseudo-random permutations: Luby-Rackoff revisited," manuscript, 1996.