Probabilistic Analysis of Elementary Randomizers

Josef Pieprzyk *

Department of Computer Science
University College
University of New South Wales
Australian Defence Force Academy
Canberra, ACT 2600, AUSTRALIA

Abstract

In the paper, elementary randomizers based on random functions and the DES structure are examined. First, it is proved that the randomizer with three different random functions produces the outputs which are independent and uniformly distributed random variables. Next, randomizers based on two different random functions are considered and it is shown that their statistical properties depend upon the order of the functions used in them. Finally, it is proved that the randomizer with a single random function gives outputs which are statistically related.

1 Introduction

Luby and Rackoff [1] introduced an elementary randomizer $\psi(f,g,h)$ based on three random functions f,g,h and the DES structure. They proved that such randomizer cannot be efficiently distinguished from a truly random permutation (function). Ohnishi [2] showed that it is possible to simplify the Luby-Rackoff randomizer to $\psi(f,f,g)$ without any significant deterioration of its quality - it cannot be distinguished from a truly random permutation as well. Schnorr [5] asked about the possibility of a further reduction of the number of random (or pseudorandom) functions to a single one. Pieprzyk [3] proved that $\psi(f,f,f,f^2)$ is indistinguishable from a truly random permutation, when f is a truly random or pseudorandom function.

The four elementary randomizers $\psi(f,g,h)$, $\psi(f,f,g)$, $\psi(g,f,f)$, and $\psi(f,f,f,f^2)$ are not distinguishable from a truly random permutation and in this paper, we will examine the statistical properties of their outputs.

*Support for this project was provided in part by TELECOM Australia under the contract number 7027 and by the Australian Research Council under the reference number A48830241.

D.W. Davies (Ed.): Advances in Cryptology - EUROCRYPT '91, LNCS 547, pp. 542-546, 1991.

2 Preliminaries

Let $I_n = \{0,1\}^n$ be the set of all 2^n binary strings of length n. For $a, b \in I_n$, $a \oplus b$ stands for bit-by-bit exclusive-or of a and b. The set of all functions from I_n to I_n is F_n, i.e., $F_n = \{f \mid f : I_n \rightarrow I_n\}$. If we have two functions $f, g \in F_n$, their composition $f \circ g$ is denoted as $f \circ g(x) = f(g(x))$ for all $x \in I_n$. For a function $f \in F_n$, we define the DES-like permutation associated with f as $D_{2n,f}(L,R) = (R, L \oplus f(R))$, where L and R are n-bit strings $(L, R \in I_n)$ and $D_{2n,f} \in F_{2n}$. Having a sequence of functions $f_1, f_2, \cdots, f_i \in F_n$, we can determine the composition of their DES-like permutations ψ and $\psi(f_1, f_2, \cdots, f_i) = D_{2n,f_i} \circ D_{2n,f_{i-1}} \circ \cdots \circ D_{2n,f_1}$ Of course, $\psi(f_1, f_2, \cdots, f_i) \in F_{2n}$.

Definition 2.1 *A random function $f \in_R F_n$ is a sequence $(f(0), f(1), \cdots, f(2^n - 1))$ of random variables, where any random variable $f(x_i)$; $x_i \in I_n$, has a uniform and independent distribution so $P[f(x_i) = x_j] = \frac{1}{2^n}$ for all $x_i, x_j \in I_n$ ($f(x_i)$ and $f(x_j)$ are independent for $i \neq j$).*

The following properties of random variables and the exclusive-or operation are exploited in the paper:

P1. Let X and Y be independent random variables. The random variable $X \oplus Y$ may be described by its conditional probabilities

$$P[X \oplus Y = z_i \mid X = x_j] = P[X \oplus Y = z_i \mid x_j] = P[Y = y_k]$$

where $y_k = z_i \oplus x_j$.

P2. Let X and Y be independent random variables and one of them, say X, be uniformly distributed. Then $X \oplus Y$ is also a uniformly distributed random variable.

P3. Let X and Y be uniformly distributed independent random variables i.e., $P[X = x_i] = P[Y = y_i] = \frac{1}{2^n}$ for all $x_i, y_i \in I_n$. Then $X \oplus Y$ is a uniformly distributed random variable that is independent from both X and Y.

3 Analysis of elementary randomizers

In this section, elementary randomizers $\psi(f,g,h)$, $\psi(f,f,g)$, $\psi(g,f,f)$, and $\psi(f,f,f,f^2)$ are analysed. First, we consider the Luby-Rackoff randomizer.

Theorem 3.1 *Given the L-R randomizer $\psi(f,g,h)$, where $f, g, h \in_R F_n$ are three different random functions, then its outputs (β, γ) are represented by two independent random variables each of the uniform probability distribution.*

Proof: The following notations are used in the proof: $X = f(R)$, $\alpha = L \oplus X$, $Y = g(\alpha)$, $\beta = Y \oplus R$, $Z = h(\beta)$, and $\gamma = Z \oplus \alpha$. The input R specifies a single random variable $X = f(R)$ in the random function f. Clearly, $P[f(R) = i] = \frac{1}{2^n}$, where $i \in I_n$ so the random variable X is uniformly distributed. The second random function g operates on values of the random variable $\alpha = X \oplus L$ (of the uniform distribution). The resulting variable $Y = g(\alpha)$ has the uniform distribution and because g is a collection of independent random variables, Y is independent from α (it is easy to check that $P[Y = i \mid \alpha = j] = P[Y = i] = \frac{1}{2^n}$ for all $i, j \in I_n$). As $\beta = Y \oplus R$ is permuted random variable Y (a deterministic transformation of Y), β is independent from α and has the uniform distribution. The application of the third random function h generates the random variable $Z = g(\beta)$ which is uniformly distributed and independent from β. As α and Z are independent and uniform, $\gamma = \alpha \oplus Z$ is the uniform random variable and independent from

both β and Z (see the property P3). Thus, the outputs (β, γ) are independent and uniformly distributed random variables.

□

Theorem 3.2 *Given the elementary randomizer $\psi(f, f, g)$, where $f, g \in_R F_n$ are two different random functions, then its outputs (β, γ) are represented by two independent random variables and γ has the uniform distribution (β does not have the uniform distribution).*

Proof: Observe that if the random variable $X = f(R) = L \oplus R$ in the first call of f, then $\alpha = X \oplus R = R$ and in the second call of f, the same value must be used. Thus $P[Y = L \oplus R \mid f(R) = L \oplus R] = 1$. Other conditional probabilities are as follows: $P[Y = i \mid f(R) = j; j \neq L \oplus R] = \frac{1}{2^n}$ for all $i, j \in I_n$. It means that

$$P[Y = L \oplus R] = \frac{2^n - 1}{2^{2n}} + \frac{1}{2^n}$$

and for all $j \neq L \oplus R$; $j \in I_n$, we have

$$P[Y = j] = \frac{2^n - 1}{2^{2n}}$$

As $\beta = Y \oplus R$, β has the following probability distribution:

$$P[\beta = i] = \begin{cases} \frac{2^n-1}{2^{2n}} + \frac{1}{2^n} & \text{if } i{=}L \\ \frac{2^n-1}{2^{2n}} & \text{otherwise} \end{cases}$$

Clearly, the application of the second random function g generates a uniform random variable $Z = g(\beta)$ which is independent from β and according to the property P2, $\gamma = Z \oplus \alpha$ is uniformly distributed and independent from β.

□

Theorem 3.3 *Given the elementary randomizer $\psi(g, f, f)$, where $f, g \in_R F_n$ are two different random functions, then its outputs (β, γ) are represented by two independent random variables and the both have the uniform probability distribution.*

Proof: R assigns a random variable $X = g(R)$ from the random function g. Clearly, X has the uniform distribution. The input L permutes values of X, and the resulting random variable $\alpha = X \oplus L$ is also uniform. Values of α are arguments of the second random function f which is independent from g. As all random variables $f(\alpha_i)$; $\alpha_i \in I_n$, are uniformly distributed, the random variable $Y = f(\alpha)$ is uniform and so is the random variable $\beta = Y \oplus R$. The probability distribution of $Z = f(\beta)$ is not uniform as it depends on $f(\alpha_i)$ and $P[Z = \alpha_i \oplus R \mid f(\alpha_i) = \alpha_i \oplus R] = 1$. In the rest of the cases (i.e., $j \neq \alpha_i \oplus R$), the probabilities are as follows: $P[Z = j \mid f(\alpha_i) \neq \alpha_i \oplus R] = \frac{1}{2^n}$. Therefore the probability distribution of Z is

$$P[Z = j] = \begin{cases} \frac{2^n-1}{2^{2n}} + \frac{1}{2^n} & \text{if } j{=}\alpha_i \oplus R \\ \frac{2^n-1}{2^{2n}} & \text{otherwise} \end{cases}$$

Obviously, α and Z are independent and according to the property P2, $\gamma = \alpha \oplus Z$ is uniformly distributed and independent from β.

□

Theorem 3.4 *Given $\psi(f, f, f, f^2)$ where $f \in_R F_n$. If the inputs $(L \neq R)$, then the outputs (γ, δ) are statistically related and their probability distributions are not uniform.*

Proof: We use the following notations: $X = f(R)$, $\alpha = L \oplus X$, $Y = f(\alpha)$, $\beta = Y \oplus R$, $Z = f(\beta)$, $\gamma = Z \oplus \alpha$ and $\delta = \beta \oplus f^2(\gamma)$. We are going to calculate conditional probabilities of variables Z and γ provided that $f(R)$ is fixed. There are two different cases: the first when $f(R) = i$ and $i \neq L \oplus R$ ($i \in I_n$), and the second when $f(R) = L \oplus R$.

1. $f(R) = i$ and $i \neq L \oplus R$. Observe that $f(R) = i$ implies that $\alpha = L \oplus f(R)$ becomes $i \oplus L$. The probability of attaining the points of Z can be found by starting from the point $\alpha = i \oplus L$ and counting all possible paths (there are 2^n of them) along with their probabilities. The probabilities are

$$\begin{aligned} P(Z = i \oplus L \oplus R \mid f(R) = i) &= P(\gamma = R \mid f(R) = i) = \frac{2^n - 2}{2^{2n}} + \frac{1}{2^n} \\ P(Z = i \mid f(R) = i) &= P(\gamma = L \mid f(R) = i) = \frac{2^n - 2}{2^{2n}} + \frac{1}{2^n} \\ P(Z = j \mid f(R) = i) &= P(\gamma = j \oplus i \oplus L \mid f(R) = i) = \frac{2^n - 2}{2^{2n}} \end{aligned}$$

   where $j \in I_n$ and is different from i and $i \oplus L \oplus R$.

2. $f(R) = L \oplus R$. It means that $\alpha = R$ and Z has a uniform distribution thus

$$P(Z = j \mid f(R) = L \oplus R) = P(\gamma = j \oplus R \mid f(R) = L \oplus R) = \frac{1}{2^n}$$

   for all $j \in I_n$.

The probability distribution of γ is as follows:

$$\begin{aligned} P(\gamma = L) &= \sum_{k \neq L \oplus R} P(\gamma = L \mid f(R) = k) P(f(R) = k) + \\ &\quad P(\gamma = L \mid f(R) = L \oplus R) P(f(R) = L \oplus R) = \\ &\quad \frac{(2^n - 1)(2^n - 2)}{2^{3n}} + \frac{1}{2^n} \\ P(\gamma = R) &= \sum_{k \neq L \oplus R} P(\gamma = R \mid f(R) = k) P(f(R) = k) + \\ &\quad P(\gamma = R \mid f(R) = L \oplus R) P(f(R) = L \oplus R) = \\ &\quad \frac{(2^n - 1)(2^n - 2)}{2^{3n}} + \frac{1}{2^n} \\ P(\gamma = j) &= \sum_{k \neq L \oplus R} P(\gamma = j \mid f(R) = k) P(f(R) = k) + \\ &\quad P(\gamma = j \mid f(R) = L \oplus R) P(f(R) = L \oplus R) = \\ &\quad \frac{(2^n - 1)(2^n - 2)}{2^{3n}} + \frac{1}{2^{2n}} \end{aligned}$$

where $j \in I_n$ and is different from L and R.

Clearly the second output variable $\delta = f^2(\gamma) \oplus \beta$ not only depends upon the random function f but also upon γ and its probability distribution is not uniform.

□

If the inputs to $\psi(f, f, f, f^2)$ are the same i.e., $L = R$, then the probability distribution of γ has a single point with bigger probability and

$$\begin{aligned} P(\gamma = R = L) &= \frac{(2^n - 1)(2^n - 2)}{2^{3n}} + \frac{2(2^n - 1)}{2^{2n}} + \frac{1}{2^n} \\ P(\gamma = j) &= \frac{(2^n - 1)(2^n - 2)}{2^{3n}} \end{aligned}$$

where $j \in I_n$ and is different from L.

4 Conclusions

We can define a perfect randomizer as the one whose outputs are statistically independent for different inputs. It also means that oracle gates of a distinguisher evaluated by a perfect randomizer are not "transparent" for the input. From the analysis, we can conclude that the composition of randomizers $\psi(f, f, f, f^2)$ does not provide a perfect randomizer. Perhaps perfect randomizers can be built from $\psi(g, f, f)$. However, the direct composition of $\psi(f, f, g)$ does not yield perfect randomizers. The composition of $\psi(f, g, h)$ creates a perfect randomizer (see [4]).

ACKNOWLEDGMENT

I would like to thank Lawrie Brown and other members of the CCSR for their assistance.

References

[1] M. Luby and Ch. Rackoff. How to construct pseudorandom permutations from pseudorandom functions. *SIAM Journal on Computing*, 17(2):373–386, April 1988.

[2] Y. Ohnishi. A study on data security. Master Thesis, Tohoku University, Japan, 1988. (in Japanese).

[3] J.P. Pieprzyk. How to construct pseudorandom permutations from single pseudorandom functions. Abstracts of EUROCRYPT'90, May 1990.

[4] J.P. Pieprzyk and R. Safavi-Naini. Randomized authentication systems. In *EUROCRYPT'91, Brighton, UK*, April 1991.

[5] C.P. Schnorr. On the construction of random number generators and random function generators. In *Proc. of Eurocrypt 88, Lecture Notes in Computer Science*, New York, 1988. Springer Verlag.