T-79.512 Cryptology: Special topics / Provable Security.

Optimal Distinguisher of Random Functions and Random Permutations.

Oleg Mürk

1 Introduction

In this paper we shall construct a distinguisher between random functions and random permutations that can make limited amount of queries. We shall also prove that its advantage is optimal amongst non-uniform randomized algorithms. At first, we introduce notation used.

Definition 1.1 For a given positive integer L we define sets Fun^L and $Perm^L$ of respectively all functions and all permutations of type $\mathbb{Z}_L \to \mathbb{Z}_L$.

Definition 1.2 A *distinguisher* $D^f(L,q)$ is a non-uniform randomized algorithm[1] that takes as input positive integers L and q, can query function f of type $\mathbb{Z}_L \to \mathbb{Z}_L$ at most q times, and returns either 0 or 1.

When L and q are evident from the context, we shall omit them and just write D^f.

Definition 1.3 For given L, q, f, and execution of distinguisher $D^f(L,q)$ the *transcript* is a sequence of pairs $(x_i, y_i) : i \in [0 \dots q-1]$ that resulted from the queries of the distinguisher, where $y_i = f(x_i)$.

Definition 1.4 We say that a transcript has a *collision*, if $\exists i,j : (x_i \neq x_j) \wedge (y_i = y_j)$.

Definition 1.5 For given L and q and two sets X and Y of functions of type $\mathbb{Z}_L \to \mathbb{Z}_L$ the *advantage* $Adv_{L,q}^{X,Y}(D^f)$ of distinguisher $D^f(L,q)$ is

$$\left|P_{f\in X}(D^f(L,q)=1) - P_{f\in Y}(D^f(L,q)=1)\right|.$$

When L and q are evident from the context we shall just write $Adv^{X,Y}(D^f)$.

[1] Non-uniform means that different Turing's Machines (TM) can be selected for different L and q.

2 Distinguisher Construction

In this section we shall describe our distinguisher construction $D_*^f(L,q)$ and compute its advantage.

Algorithm 2.1 *The distinguisher $D_*^f(L,q)$ acts as follows:*

1. *Query function f at the points $0 \ldots q-1$.*
2. *Return 0 if the transcript has collisions and 1 otherwise.*

Theorem 2.2 *For given L and q it holds:*

$$Adv^{Fun^L,Perm^L}(D_*^f) = 1 - \frac{(L)_q}{L^q}$$

Proof. It is easy to see that $P_{f \in Perm^L}(D^f = 1) = 1$ and $P_{f \in Fun^L}(D^f = 1) = \frac{(L)_q L^{L-q}}{L^L} = \frac{(L)_q}{L^q}$. ■

3 Optimality of the Distinguisher Construction

In this section we shall prove an upper bound for the advantage of distinguisher between random functions and random permutations. It will turn out that $D_*^f(L,q)$ has the same advantage. Consequently our construction is optimal.

Theorem 3.1 *For given L and q and any distinguisher D^f it holds:*

$$Adv^{Fun^L,Perm^L}(D^f) \leq 1 - \frac{(L)_q}{L^q}$$

Proof. Let's fix L and q. Now $D^f(L,q)$ is just a randomized TM. We shall prove that the bound holds for any deterministic TM. From this we can directly conclude that the same bound holds for a randomized TM. That is because random bits are selected independently of the function f. We shall also limit ourselves to distinguishers that make exactly q queries and do not ask queries for x_k if it is already present in the partial transcript $(x_i, y_i) : i \in [0 \ldots k-1]$. Any distinguisher can be easily transformed to this form. We shall call the resulting deterministic TM A^f. Let's denote the transcript of A^f by $\sigma(A^f)$.

For given A^f we can split set Fun^L into two disjoint subsets:

$$\begin{aligned} CFun^L &= \{f : f \in Fun^L, \sigma(A^f) \text{ has a collision}\}, \\ NCFun^L &= Fun^L \setminus CFun^L. \end{aligned}$$

Note that $Perm^L \subseteq NCFun^L$. From distinguisher A^f it is easy to construct distinguishers A_0^f and A_1^f that return respectively 0 and 1 when $f \in CFun^L$. That is because collisions are easy to detect. It is easy to see that for both i the probability $P_{f \in Perm^L}(A_i^f = 1)$ is equal to $P_{f \in Perm^L}(A^f = 1)$. Also, for some i

$$P_{f \in Fun^L}(A_i^f = 1) \leq P_{f \in Fun^L}(A^f = 1)$$

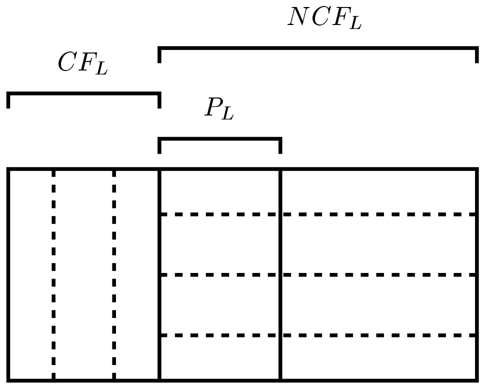

Figure 1

and for the other

$$P_{f\in Fun^L}(A^f_{1-i}=1)\geq P_{f\in Fun^L}(A^f=1).$$

Consequently one of the distinguishers A^f_i has at least as big advantage as A^f. If that is A^f_1 we can invert its inputs without changing the advantage. Consequently, we can limit ourselves to distinguishers that always return 0 for $f\in CFun^L$.

Now, let's define an equivalence relation $f\sim g \Leftrightarrow \sigma(A^f)=\sigma(A^g)$. Let's denote a set of equivalence classes on a set of functions X according to the relation $\sim$ by X/σ. It quite easy to see that:

1. Any distinguisher returns the same output for every element of an equivalence class from either Fun^L/σ or $Perm^L/\sigma$.

2. Every element of Fun^L/σ has size L^{L-q}.

3. Every element of $Perm^L/\sigma$ has size $(L-q)!$.

4. $Fun^L/\sigma = CFun^L/\sigma \cup NCFun^L/\sigma$.

5. There exists a bijection $q: Perm^L/\sigma \to NCFun^L/\sigma$ such that $\forall x\in Perm^L/\sigma : x\subseteq q(x)$.

An illustration of the situation is given on Figure 1. Here dashed lines signify equivalence class boundaries. From these facts it is easy to conclude that

$$\#\{A^f=1: f\in NCFun^L\} = \frac{L^{L-q}}{(L-q)!}\cdot \#\{A^f=1: f\in Perm^L\}.$$

Now we can compute the advantage:

$$\begin{aligned}
Adv^{Fun^L,Perm^L}(A^f) &= \left|P_{f\in Fun^L}(A^f=1) - P_{f\in Perm^L}(A^f=1)\right| \\
&= \left|\tfrac{\#\{A^f=1:f\in Fun^L\}}{\#Fun^L} - \tfrac{\#\{A^f=1:f\in Perm^L\}}{\#Perm^L}\right| \\
&= \left|\tfrac{\#\{A^f=1:f\in NCFun^L\}}{L^L} - \tfrac{\#\{A^f=1:f\in Perm^L\}}{L!}\right| \\
&= \left|\#\{A^f=1:f\in Perm^L\}\left(\tfrac{\frac{L^{L-q}}{(L-q)!}}{L^L} - \tfrac{1}{L!}\right)\right|
\end{aligned}$$

The largest possible value of $\#\{A^f = 1 : f \in Perm^L\}$ is $L!$. Consequently the upper bound for the advantage is

$$\left|L!\left(\frac{\frac{L^{L-q}}{(L-q)!}}{L^L} - \frac{1}{L!}\right)\right| = \left|\frac{(L)_q}{L^q} - 1\right| = 1 - \frac{(L)_q}{L^q}.$$

This finishes the proof. ■

4 Remarks

With relatively small technical effort it is possible to prove that distinguiser $D_*^f(L,q)$ is also optimal when queries to f^{-1} can be made.

Careful reader has probably already noticed that all this time we have been dealing with the "birthday" attack and $1-\frac{(L)_q}{L^q}$ is just a birthday attack success probability $C(L,q)$.

Curious reader can find more-or-less the same construction in [GB01]: subsections 5.9 "Birthday Attack" and 5.10 "PRFs versus PRPs". The statement they are essentially proving in Proposition 5.17 is in our notation

$$Adv^{Fun^L,E^L}(D^f) \leq Adv^{Fun^L,Perm^L}(D^f) + Adv^{Perm^L,E^L}(D^f)$$

$$Adv^{Fun^L,Perm^L}(D^f) \leq C(L,q),$$

where E is an arbitrary subset of $Perm^L$ (usually referred to as pseudorandom permutation). The last formula of the proof of this proposition is essentially a proof of our Theorem 3.1, although they implicitly assume that

$$P[B^g = 1 \mid D] = P[Exp_{E,B}^{\text{prp-cpa-0}} = 1],$$

which is equivalent in our notation to

$$P_{f\in NCFun^L}(B^f = 1) = P_{f\in Perm^L}(B^f = 1).$$

This is the claim, on which most effort of our proof is spent. Similar claim can be found in [BDJR97] in Proposition 8:

$$Adv^{Fun^L,E^L}(D^f) \leq q^2 2^{-L-1} + Adv^{Perm^L,E^L}(D^f).$$

References

[GB01] S. Goldwasser and M. Bellare. Lecture Notes on Cryptography.

[BDJR97] M. Bellare, A. Desai, E. Jokipii, M. Rogaway. A Concrete Security Treatment of Symmetric Encryption: Analysis of the DES Modes of Operation.