J Cryptol (2018) 31:162–171
https://doi.org/10.1007/s00145-017-9253-0

Journal of
CRYPTOLOGY

How Many Queries are Needed to Distinguish a Truncated Random Permutation from a Random Function?

Shoni Gilboa
The Open University of Israel, Ra'anana, Israel

Shay Gueron
University of Haifa, Haifa, Israel
shay@math.haifa.ac.il

Amazon Web Services, Seattle, WA, USA

Ben Morris
UC Davis, Davis, CA, USA

Communicated by Philip Rogaway.

Received 30 September 2014 / Revised 30 January 2017
Online publication 7 April 2017

Abstract. An oracle chooses a function f from the set of n bits strings to itself, which is either a randomly chosen permutation or a randomly chosen function. When queried by an n-bit string w, the oracle computes $f(w)$, truncates the m last bits, and returns only the first $n-m$ bits of $f(w)$. How many queries does a querying adversary need to submit in order to distinguish the truncated permutation from the (truncated) function? In Hall et al. (Building PRFs from PRPs, Springer, Berlin, 1998) showed an algorithm for determining (with high probability) whether or not f is a permutation, using $O(2^{\frac{m+n}{2}})$ queries. They also showed that if $m < n/7$, a smaller number of queries will not suffice. For $m > n/7$, their method gives a weaker bound. In this note, we first show how a modification of the approximation method used by Hall et al. can solve the problem completely. It extends the result to practically any m, showing that $\Omega(2^{\frac{m+n}{2}})$ queries are needed to get a non-negligible distinguishing advantage. However, more surprisingly, a better bound for the distinguishing advantage, which we can write, in a simplified form, as $O\left(\min\left\{\frac{q^2}{2^n}, \frac{q}{2^{\frac{n+m}{2}}}, 1\right\}\right)$, can be obtained from a result of Stam published, in a different context, already in 1978. We also show that, at least in some cases, this bound is tight.

Keywords. Pseudo-random permutations, Pseudo-random functions, Advantage.

1. Introduction

Distinguishing a randomly chosen permutation from a random function is a combinatorial problem which is fundamental in cryptology. A few examples where this problem plays an important role are the security analysis of block ciphers, hash, and MAC schemes.

One formulation of this problem is the following. An oracle chooses a function $F : \{0,1\}^n \to \{0,1\}^n$, which is either a randomly (uniformly) chosen permutation of $\{0,1\}^n$ or a randomly (uniformly) chosen function from $\{0,1\}^n$ to $\{0,1\}^n$. An adversary selects a "querying and guessing" algorithm. He first uses it to submit q (adaptive) queries to the oracle, and the oracle responds with $F(w)$ to the query $w \in \{0,1\}^n$. After collecting the q responses, the adversary uses his algorithm to guess whether or not F is a permutation. The quality of such an algorithm (in the cryptographic context) is the ability to distinguish between the two cases (rather than successfully guessing which one it is). It is measured by the difference between the probability that the algorithm outputs a certain answer, given that the oracle chose a permutation, and the probability that the algorithm outputs the same answer, given that the oracle chose a function. This difference is called the "advantage" of the algorithm. We are interested in estimating *Adv*, which is the maximal advantage of the adversary, over all possible algorithms, as a function of a budget of q queries.

The well-known (folklore) answer to this problem is based on the simple "collision test" and the Birthday Problem:

$$\mathrm{Adv} = 1 - \left(1 - \frac{1}{2^n}\right)\left(1 - \frac{2}{2^n}\right)\ldots\left(1 - \frac{q-1}{2^n}\right).$$

Since for every $1 \le k \le q-1$

$$1 - \frac{q}{2^n} \le \left(1 - \frac{k}{2^n}\right)\left(1 - \frac{q-k}{2^n}\right) \le \left(1 - \frac{q}{2^{n+1}}\right)^2,$$

we get, for $q \le 2^n$, that

$$1 - e^{-\frac{q(q-1)}{2^{n+1}}} \le 1 - \left(1 - \frac{q}{2^{n+1}}\right)^{q-1} \le \mathrm{Adv} \le 1 - \left(1 - \frac{q}{2^n}\right)^{\frac{q-1}{2}} \le \frac{q(q-1)}{2^{n+1}}. \quad (1)$$

This result implies that the number of queries required to distinguish a random permutation from a random function, with success probability significantly larger than, say, $\frac{1}{2}$, is $\Theta(2^{\frac{n}{2}})$. We now consider the following generalization of this problem:

Problem 1. (*Distinguishing a truncated permutation*) Let $0 \le m < n$ be integers. An oracle chooses $c \in \{0,1\}$. If $c = 1$, it picks a permutation p of $\{0,1\}^n$ uniformly at random, and if $c = 0$, it picks a function $f : \{0,1\}^n \to \{0,1\}^n$ uniformly at random. An adversary is allowed to submit queries $w \in \{0,1\}^n$ to the oracle. The oracle computes $\alpha = p(w)$ (if $c = 1$) or $\alpha = f(w)$ (if $c = 0$), truncates (with no loss of generality) the last m bits from α, and replies with the remaining $(n-m)$ bits. The adversary has a

budget of q (adaptive) queries, and after exhausting this budget, is expected to guess c. *How many queries does the adversary need in order to gain non-negligible advantage?* Specifically, we seek $q_{\frac{1}{2}}(n,m) = \min\{q \mid \mathbf{Adv}_{n,m}(q) \geq \frac{1}{2}\}$ as a function of m and n.

2. So, How Many Queries are Really Needed?

The Birthday bound (folklore) We start with remarking that the classical "Birthday" bound $q_{\frac{1}{2}}(n,m) = \Omega(2^{n/2})$ is obviously valid as a bound for the adversary's advantage in Problem 1. In fact, any algorithm that the adversary can use with the truncated replies of $(n-m)$ bits from $f(w)$ can also be used by the adversary who sees the full $f(w)$ (he can ignore m bits and apply the same algorithm).

Of course, we are looking for a better upper bound that would reflect the fact that the adversary receives less information when $f(w)$ is truncated. We have the following bounds for Problem 1.

Hall et al. [5] Problem 1 was studied by Hall et al. [5]. The authors showed an algorithm that gives a non-negligible distinguishing advantage using $q = O(2^{(n+m)/2})$ queries (for any m). They also proved the following upper bound:

$$\mathbf{Adv}_{n,m}(q) \leq 5\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{\frac{2}{3}} + \frac{1}{2}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{3}\frac{1}{2^{\frac{n-7m}{2}}}. \tag{2}$$

For $m \leq n/7$ the bound in (2) implies that $q_{\frac{1}{2}}(n,m) = \Omega(2^{\frac{m+n}{2}})$. However, for larger values of m, the bound on $q_{\frac{1}{2}}(n,m)$ that is offered by (2) deteriorates, and becomes (already for $m > n/4$)worse than the trivial "Birthday" bound $q_{\frac{1}{2}}(n,m) = \Omega(2^{n/2})$.

Hall et al. [5] conjectured that $\Omega(2^{\frac{m+n}{2}})$ queries are needed in order to get a non-negligible advantage, in the general case.

Bellare and Impagliazzo [1] Theorem 4.2 in [1] states that

$$\mathbf{Adv}_{n,m}(q) = O(n)\frac{q}{2^{\frac{n+m}{2}}} \tag{3}$$

whenever $2^{n-m} < q < 2^{\frac{n+m}{2}}$.

This implies that $q_{\frac{1}{2}}(n,m) = \Omega(\frac{1}{n}2^{\frac{m+n}{2}})$ for $m > \frac{1}{3}n + \frac{2}{3}\log_2 n + \Omega(1)$. We point out that it is hard to extract an upper bound for $\mathbf{Adv}_{n,m}$ from [1], in a form that can be directly compared to the other approximations that are discussed here.

Gilboa and Gueron [4] The method used to show (2) can be pushed to prove the conjecture in [5] for (almost) every m. In particular, it can be shown that if $m \leq n/3$ then

$$\mathbf{Adv}_{n,m}(q) \leq 2\sqrt[3]{2}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{\frac{2}{3}} + \frac{2\sqrt{2}}{\sqrt{3}}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{\frac{3}{2}} + \left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{2}, \tag{4}$$

and if $\frac{n}{3} < m \leq n - 4 - \log_2 n$ then

$$\mathbf{Adv}_{n,m}(q) \leq 3\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{\frac{2}{3}} + 2\left(\frac{q}{2^{\frac{n+m}{2}}}\right) + 5\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{2} + \frac{1}{2}\left(\frac{2q}{2^{\frac{n+m}{2}}}\right)^{\frac{n}{n-m}}. \quad (5)$$

This implies that $q_{\frac{1}{2}}(n, m) = \Omega(2^{\frac{m+n}{2}})$ for any $0 \leq m \leq n - 4 - \log_2(n)$.
Stam [9] Surprisingly, it turns out that Problem 1 was solved 20 years before Hall et al. [5], in a different context. The bound

$$\mathbf{Adv}_{n,m}(q) \leq \frac{1}{2}\sqrt{\frac{(2^{n-m}-1)q(q-1)}{(2^n-1)(2^n-(q-1)}} \leq \frac{1}{2\sqrt{1-\frac{q-1}{2^n}}} \cdot \frac{q}{2^{\frac{n+m}{2}}}, \quad (6)$$

which is valid for all $0 \leq m < n$, follows directly from a result of Stam [9, Theorem 2.3]. Note that if $q \leq \frac{3}{4}2^n$ then (6) can be simplified to the very handy form

$$\mathbf{Adv}_{n,m}(q) \leq \frac{q}{2^{\frac{m+n}{2}}}. \quad (7)$$

This implies that $q_{\frac{1}{2}}(n, m) = \Omega(2^{\frac{m+n}{2}})$ for any $0 \leq m < n$, confirming the conjecture of [5] in all generality (20 years before the conjecture was raised).

Remark 1. The bound (6) is tighter than all the bounds mentioned above, with one exception: the elementary upper bound (1) is better than (6) for $q \leq 2^{\frac{n-m}{2}}$.

3. Different Methods Give Different Bounds

It is interesting to see how different approaches yield different bounds for Problem 1. To this end, we first define some notations.

For fixed $m < n$ and $q \leq 2^n$ denote $\Omega_q := \left(\{0,1\}^{n-m}\right)^q$. We view Ω_q as the set of all possible sequences of replies that can be given by the oracle (to the adversary's q queries).
For any $j \geq 2$, $\omega \in \Omega$ let

$$\mathrm{col}_j(\omega) = \#\{1 \leq i_1 < i_2 < \ldots < i_j \leq q \mid \omega_{i_1} = \omega_{i_2} = \ldots = \omega_{i_j}\}$$

For $\omega = (w_1, w_2, \ldots, w_q) \in \Omega$ and $1 \leq r \leq q$, let

$$V_r(\omega) := \{(x_1, x_2, \ldots, x_q) \in \Omega \mid \forall 1 \leq i \leq r : x_i = w_i\}$$

be the set of sequences of replies that are the same as ω up to the r-th query.

For $\omega \in \Omega$ let $\mathrm{Pr}_{\mathrm{perm}}(\omega)$ and $\mathrm{Pr}_{\mathrm{func}}(\omega)$ be the probabilities that ω is the actual sequence of replies that the oracle gives to the adversary's q queries, in the case the oracle chose a random permutation or a random function, respectively.

For $1 \leq r \leq q$, let

$$Q^{(r)}_{\text{perm}}(\omega) = \frac{\Pr_{\text{perm}}(V_r(\omega))}{\Pr_{\text{perm}}(V_{r-1}(\omega))}, \quad Q^{(r)}_{\text{func}}(\omega) = \frac{\Pr_{\text{func}}(V_r(\omega))}{\Pr_{\text{func}}(V_{r-1}(\omega))}.$$

Note that

$$\Pr_{\text{perm}}(\omega) = \prod_{r=1}^{q} Q^{(r)}_{\text{perm}}(\omega), \quad \Pr_{\text{func}}(\omega) = \prod_{r=1}^{q} Q^{(r)}_{\text{func}}(\omega).$$

3.1. *The Proof Method of Hall et al*

The proof of (2) uses the general bound

$$\begin{aligned}\mathbf{Adv}_{n,m}(q) &\leq \max_{\omega\in S}\left|\frac{\Pr_{\text{perm}}(\{\omega\})}{\Pr_{\text{func}}(\{\omega\})}-1\right| + \max\left\{\Pr_{\text{func}}(\overline{S}), \Pr_{\text{perm}}(\overline{S})\right\} \leq \\ &\leq 2\max_{\omega\in S}\left|\frac{\Pr_{\text{perm}}(\{\omega\})}{\Pr_{\text{func}}(\{\omega\})}-1\right| + \Pr_{\text{func}}(\overline{S}). \end{aligned} \tag{8}$$

that holds for any $S \subseteq \Omega$. It is applied to the set

$$S := \left\{\omega \in \Omega : \left|\text{col}_2(\omega) - \binom{q}{2}\frac{1}{2^{n-m}}\right| \leq c_1 \frac{q}{2^{\frac{n-m}{2}}}\ ,\ \text{col}_3(\omega) = 0\right\}.$$

The expression $\max_{\omega\in S}\left|\frac{\Pr_{\text{perm}}(\{\omega\})}{\Pr_{\text{func}}(\{\omega\})}-1\right|$ is bounded by direct computations. The expression $\Pr_{\text{func}}(\overline{S})$ is bounded by combining the Union Bound and the Chebyshev inequality. Finally, c_1 is chosen to minimize the resulting bounds.

3.2. *The Proof Method of Gilboa and Gueron*

To get (4) (for $m \leq n/3$), we can apply the slightly better (than (8)) bound

$$\begin{aligned}\mathbf{Adv}_{n,m}(q) &\leq \frac{1}{2}\max_{\omega\in S}\left|\frac{\Pr_{\text{perm}}(\{\omega\})}{\Pr_{\text{func}}(\{\omega\})}-1\right| + \frac{1}{2}\left(\Pr_{\text{func}}(\overline{S}) + \Pr_{\text{perm}}(\overline{S})\right) \leq \\ &\leq \max_{\omega\in S}\left|\frac{\Pr_{\text{perm}}(\{\omega\})}{\Pr_{\text{func}}(\{\omega\})}-1\right| + \min\left\{\Pr_{\text{func}}(\overline{S}), \Pr_{\text{perm}}(\overline{S})\right\} \end{aligned} \tag{9}$$

to the set

$$S := \left\{\omega \in \Omega : \left|\text{col}_2(\omega) - \binom{q}{2}\frac{1}{2^{n-m}}\right| \leq c_2 \frac{q^{2/3}2^{2m/3}}{2^{n/3}}\ ,\ \text{col}_3(\omega) \leq c_3\frac{q^{3/2}}{2^n}\right\}$$

Here, c_2, c_3 are chosen to minimize the bound. Again, $\max_{\omega\in S}\left|\frac{\Pr_{\text{perm}}(\{\omega\})}{\Pr_{\text{func}}(\{\omega\})}-1\right|$ is bounded by direct (elaborate) computation, and $\Pr_{\text{func}}(\overline{S})$ is bounded by combining (via the Union Bound) the Chebyshev inequality and the Markov inequality.

The bound (5) (for $n/3 < m \leq n - 4 - \log_2 n$) follows similarly by examining the set

$$S := \left\{ \omega \in \Omega : \left| \mathrm{col}_{j+1}(\omega) - \binom{q}{j+1} \frac{1}{2^{j(n-m)}} \right| \leq \alpha_j \ \forall 1 \leq j \leq t-1 \,,\ \mathrm{col}_{t+1}(\omega) \leq \beta \right\}$$

for $t := \left\lceil \frac{n+m}{n-m} \right\rceil$ and $\alpha_1, \ldots, \alpha_{t-1}, \beta$ which are chosen to optimize the bound.

3.3. *The Proof Method of Bellare and Impagliazzo*

Bellare and Impagliazzo also used (9), for the set S of all $\omega \in \Omega$ satisfying (for suitable δ and λ):

1. For any $1 \leq r \leq q$,

$$\left| \log \frac{Q^{(r)}_{\mathrm{perm}}(\omega)}{Q^{(r)}_{\mathrm{func}}\omega)} \right| \leq \frac{3\delta}{2}$$

2. For any $1 \leq r \leq q$,

$$\left| \sum_{x \in V_{r-1}(\omega)} \frac{\mathrm{Pr}_{\mathrm{func}}(x)}{\mathrm{Pr}_{\mathrm{func}}(V_{r-1}(\omega))} \log \frac{Q^{(r)}_{\mathrm{perm}}(x)}{Q^{(r)}_{\mathrm{func}}(x)} \right| \leq \frac{\delta^2}{2},$$

3.

$$\left| \log \frac{\mathrm{Pr}_{\mathrm{perm}}(\omega)}{\mathrm{Pr}_{\mathrm{func}}\omega)} - \sum_{r=1}^{q} \sum_{x \in V_{r-1}(\omega)} \frac{\mathrm{Pr}_{\mathrm{func}}(x)}{\mathrm{Pr}_{\mathrm{func}}(V_{r-1}(\omega))} \log \frac{Q^{(r)}_{\mathrm{perm}}(x)}{Q^{(r)}_{\mathrm{func}}(x)} \right| \leq \frac{\delta(\delta+3)\lambda\sqrt{q}}{2}.$$

The expression $\mathrm{Pr}_{\mathrm{func}}(\overline{S})$ is bounded by combining the Azuma inequality and the observation that for any $1 \leq r \leq q$,

$$0 \geq \sum_{\omega \in \Omega} Q^{(r)}_{\mathrm{func}}(\omega) \log \frac{Q^{(r)}_{\mathrm{perm}}(\omega)}{Q^{(r)}_{\mathrm{func}}(\omega)} \geq -\frac{1}{2} \left(\max_{\omega \in S} \left| \frac{Q^{(r)}_{\mathrm{perm}}(\omega)}{Q^{(r)}_{\mathrm{func}}(\omega)} - 1 \right| \right)^2,$$

3.4. *The Proof Method of Stam*

Stam's approach observes that by Pinsker's inequality [8] [1] we have

[1] The inequality as used in (10) was established independently by Csiszár [3], Kemperman [6], and Kullback [7]. Pinsker proved the inequality with a worse constant.

$$\mathbf{Adv}_{n,m}(q) \le \frac{1}{2}\sum_{\omega\in\Omega}\left|\Pr_{\text{perm}}(\omega)-\Pr_{\text{func}}(\omega)\right| \le$$
$$\le \sqrt{\frac{1}{2}\sum_{\omega\in\Omega}\Pr_{\text{perm}}(\omega)\log\frac{\Pr_{\text{perm}}(\omega)}{\Pr_{\text{func}}(\omega)}}. \tag{10}$$

He then uses the decomposition

$$\sum_{\omega\in\Omega}\Pr_{\text{perm}}(\omega)\log\frac{\Pr_{\text{perm}}(\omega)}{\Pr_{\text{func}}(\omega)} = \sum_{r=1}^{q}\sum_{\omega\in\Omega}\Pr_{\text{perm}}(V_{r-1}(\omega))Q^{(r)}_{perm}(\omega)\log\frac{Q^{(r)}_{perm}(\omega)}{Q^{(r)}_{func}(\omega)},$$

direct (exact) computations, and the concavity of the log function.

4. Stam's Bound is Sometimes Sharp

In the case $m = n-1$ (i.e., the oracle returns only 1 bit), (6) gives

$$\mathbf{Adv}_{n,n-1}(q) \le \frac{1}{2}\sqrt{\frac{q(q-1)}{(2^n-1)(2^n-(q-1)}} \le \frac{1}{\sqrt{2-\frac{q-1}{2^{n-1}}}}\cdot\frac{q}{2^n}.$$

In this section, we show that this bound is essentially sharp.

With no loss of generality, we may assume q is even and $q \le \frac{1}{2}2^n$. We define the following adversarial algorithm.

Algorithm 1. Collect the answers (which are, in this case, just bits) of q arbitrary queries.
Compute the difference Δ between the number of 0's and 1's.
If $\Delta \le \sqrt{q}/2$, guess that the oracle was using a truncated random permutation. Otherwise, guess that the oracle was using a random function.

The advantage of Algorithm 1 is

$$\sum_{|k-(q-k)|<\sqrt{q}/2}\binom{q}{k}\left(\frac{\prod_{i=1}^{k}(2^{n-1}-(i-1))\cdot\prod_{i=1}^{q-k}(2^{n-1}-(i-1))}{\prod_{i=1}^{q}(2^n-(i-1))}-\frac{1}{2^q}\right) =$$
$$= \sum_{|k-(q-k)|<\sqrt{q}/2}\binom{q}{k}\frac{1}{2^q}\left(\frac{\prod_{i=1}^{k}(2^n-2(i-1))\cdot\prod_{i=1}^{q-k}(2^n-2(i-1))}{\prod_{i=1}^{q}(2^n-(i-1))}-1\right)$$

We show that

$$\binom{q}{k}\frac{1}{2^q} \ge \frac{1}{2\sqrt{q}}, \tag{11}$$

$$p_k := \frac{\prod_{i=1}^{k}(2^n - 2(i-1)) \cdot \prod_{i=1}^{q-k}(2^n - 2(i-1))}{\prod_{i=1}^{q}(2^n - (i-1))} > 1 + \frac{q/2}{2^n} \tag{12}$$

for any k such that $|k - (q-k)| < \sqrt{q}/2$. From this, we can conclude that

$$\mathbf{Adv}_{n,n-1}(q) > \sqrt{q}\frac{1}{2\sqrt{q}}\frac{q/2}{2^n} = \frac{q/4}{2^n}.$$

First, note that for $k = q/2$.

$$\binom{q}{q/2}\frac{1}{2^q} = \frac{1}{2}\prod_{i=2}^{q/2}\frac{2i-1}{2i} \geq \frac{1}{2}\prod_{i=2}^{q/2}\frac{\sqrt{i-1}}{\sqrt{i}} = \frac{1}{\sqrt{2q}}, \tag{13}$$

$$p_{q/2} = \prod_{i=1}^{q/2}\left(1 + \frac{1}{2^n - (2i-1)}\right) \geq \left(1 + \frac{1}{\frac{1}{2}2^n}\right)^{q/2} \geq 1 + \frac{q}{2^n}. \tag{14}$$

Since for any $0 \leq j < q/2$

$$\frac{\binom{q}{j}}{\binom{q}{j+1}} = 1 - \frac{q-2j-1}{q-j} > 1 - \frac{2(q-2j-1)}{q},$$
$$\frac{p_j}{p_{j+1}} = 1 - \frac{2(q-2j-1)}{2^n - 2j} \geq 1 - \frac{4(q-2j-1)}{2^n},$$

we get that for any $\frac{q}{2} - \frac{\sqrt{q}}{4} \leq k < \frac{q}{2}$

$$\begin{aligned}
\frac{\binom{q}{k}}{\binom{q}{q/2}} &= \prod_{i=k}^{\frac{q}{2}-1}\frac{\binom{q}{j}}{\binom{q}{j+1}} \geq \prod_{j=k}^{\frac{q}{2}-1}\left(1 - \frac{2(q-2j-1)}{q}\right) \\
&\geq 1 - \frac{2\sum_{j=k}^{\frac{q}{2}-1}(q-2j-1)}{q} = \\
&= 1 - \frac{(q-2k)^2}{2q} \geq \frac{7}{8}, \\
\frac{p_k}{p_{q/2}} &= \prod_{i=k}^{\frac{q}{2}-1}\frac{p_j}{p_{j+1}} \geq \prod_{j=k}^{\frac{q}{2}-1}\left(1 - \frac{4(q-2j-1)}{2^n}\right) \\
&\geq 1 - \frac{4\sum_{j=k}^{\frac{q}{2}-1}(q-2j-1)}{2^n} = \\
&= 1 - \frac{(q-2k)^2}{2^n} \geq 1 - \frac{q/4}{2^n}.
\end{aligned}$$

Now, using (13) and (14) we get

$$\binom{q}{k}\frac{1}{2^q} = \frac{\binom{q}{k}}{\binom{q}{q/2}}\binom{q}{q/2}\frac{1}{2^q} \geq \frac{7}{8}\cdot\frac{1}{\sqrt{2q}} > \frac{1}{2\sqrt{q}},$$

$$p_k = \frac{p_k}{p_{q/2}}p_{q/2} \geq \left(1-\frac{q/4}{2^n}\right)\left(1+\frac{q}{2^n}\right) > 1+\frac{q/2}{2^n}.$$

The proof of (11) and (12) for $\frac{q}{2} < k \leq \frac{q}{2}+\frac{\sqrt{q}}{4}$ is similar.

5. An Open Problem

By combining (1), (6), and the trivial bound 1, we can conclude that the best known bound for Problem 1 is

$$\mathbf{Adv}_{n,m}(q) \leq \begin{cases} \frac{q(q-1)}{2^{n+1}} & q < (1+o(1))\,2^{\frac{n-m}{2}} \\ \frac{1}{2}\sqrt{\frac{(2^{n-m}-1)q(q-1)}{(2^n-1)(2^n-(q-1)}} & (1+o(1))\,2^{\frac{n-m}{2}} \leq q \leq (2+o(1))\,2^{\frac{n+m}{2}} \\ 1 & (2+o(1))\,2^{\frac{n+m}{2}} < q, \end{cases} \quad (15)$$

and in a simpler form:

$$\mathbf{Adv}_{n,m}(q) = O\left(\min\left\{\frac{q^2}{2^n}, \frac{q}{2^{\frac{n+m}{2}}}, 1\right\}\right). \quad (16)$$

Figure 1 shows the graphs of the base 2 logarithm of $\frac{q^2}{2^n}$ and $\frac{q}{2^{\frac{n+m}{2}}}$ as a function of q, for different ranges of q, illustrating the crossover point at $q = 2^{\frac{n-m}{2}}$.

By the lower bound in (1), we know that the bound in (16) is essentially sharp for $m = 0$. By our proof in Sect. 4, we know that the bound in (16) is essentially sharp

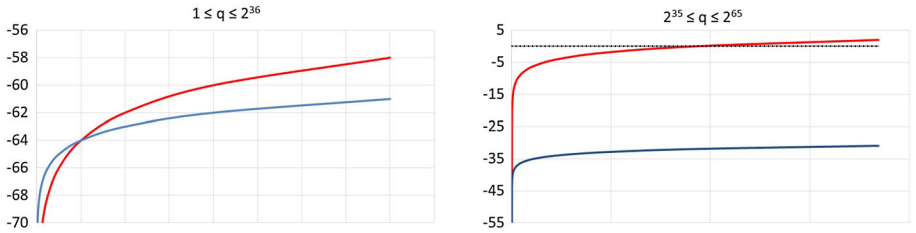

Fig. 1. Base 2 logarithm of $\frac{q^2}{2^n}$ (*red line*) and $\frac{q}{2^{\frac{n+m}{2}}}$ (*blue lines*), which appear in the upper bound 16. Here, $n = 128$, $m = n/2 = 64$, and the functions are plotted for low (*left*) and high (*right*) ranges of q (the scale of the horizontal axis is logarithmic). The value at $q = 2^{32}$) is the crossover point, were the "linear" term (*blue line*) provides the better upper bound than the "quadratic" term (*red line*). Note that the latter term becomes worse than the trivial bound ($\log_2(1) = 0$) at $q = 2^{64}$ (Color figure online).

$m = n-1$. The natural question that remains open is whether the bound (16) is essentially sharp for all $0 \leq m < n$.

Added in proof Since the paper was submitted, the two first authors managed to solve the above question, and to prove that the bound (16) is essentially tight. See Ref. [2].

References

[1] M. Bellare, R. Impagliazzo, A tool for obtaining tighter security analyses of pseudorandom function based constructions, with applications to PRP to PRF conversion. ePrint 1999/024. http://eprint.iacr.org/1999/024.

[2] I. Csiszár, Information-type measures of difference of probability distributions and indirect observations, Stud. Sci. Math. Hung. **2** (1967), 299–318.

[3] S. Gilboa, S. Gueron, Distinguishing a truncated random permutation from a random function. ePrint 2015/773. http://eprint.iacr.org/2015/773.

[4] S. Gilboa, S. Gueron, The Advantage of Truncated Permutations. Available at arXiv:1610.02518.

[5] C. Hall, D. Wagner, J. Kelsey, B. Schneier, Building PRFs from PRPs, in: *Proceedings of CRYPTO-98: Advances in Cryptography* (Springer, Berlin, 1998), pp. 370–389.

[6] J. H. B. Kemperman, On the optimum rate of transmitting information, in *Probability and Information Theory (Proc. Int. Symp., McMaster Univ., Hamilton, Ont.)*, (Springer, Berlin, 1968), pp. 126–169.

[7] S. Kullback, A lower bound for discrimination information in terms of variation, IEEE Trans. Inf. Theory 13 (1967), 126–127.

[8] M. S. Pinsker, *Information and informational stability of random variables and processes* (Russian), Problemy Peredači Informacii, Vyp. 7, Akad. Nauk SSSR, Moscow (1960).

[9] A. J. Stam, Distance between sampling with and without replacement, Stat. Neerlandica **32** (1978), no. 2, 81–91.