Balanced permutations Even-Mansour ciphers

Shoni Gilboa[1] and Shay Gueron[2,3]

[1] The Open University of Israel, Raanana 43107, Israel
[2] University of Haifa, Israel
[3] Intel Corporation, Israel Development Center, Israel

July 21, 2014

Abstract. The r-rounds Even-Mansour block cipher uses r public permutations of $\{0,1\}^n$ and $r+1$ secret keys. An attack on this construction was described in [6], for $r = 2, 3$. Although this attack is only marginally better than brute force, it is based on an interesting observation (due to [10]): for a "typical" permutation P, the distribution of $P(x) \oplus x$ is not uniform. To address this, and other potential threats that might stem from this observation in this (or other) context, we introduce the notion of a "balanced permutation" for which the distribution of $P(x) \oplus x$ is uniform, and show how to generate families of balanced permutations from the Feistel construction. This allows us to define a $2n$-bit block cipher from the 2-rounds Even-Mansour scheme. The cipher uses public balanced permutations of $\{0,1\}^{2n}$, which are based on two public permutations of $\{0,1\}^n$. By construction, this cipher is immune against attacks that rely on the non-uniform behavior of $P(x) \oplus x$. We prove that this cipher is indistinguishable from a random permutation of $\{0,1\}^{2n}$, for any adversary who has oracle access to the public permutations and to an encryption/decryption oracle, as long as the number of queries is $o(2^{n/2})$. As a practical example, we discuss the properties and the performance of a 256-bit block cipher that is based on AES.

Keywords: Even-Mansour, block-cipher, Feistel rounds

1 Introduction

The Even-Mansour (EM) block cipher [7] uses a randomly chosen (public) permutation P of $\{0,1\}^n$, two secret keys $k_1, k_2 \in \{0,1\}^n$, and encrypts an n-bit plaintext m by the transformation $EM_{P;k_1,k_2}(m) = P(m \oplus k_1) \oplus k_2$ (the decryption transformation is obvious). The interesting feature of the EM cipher is that confidentiality is achieved even if the permutation P is made public. An adversary needs to make at least $2^{n/2}$ oracle queries before he can decrypt a new message with high success probability (the model adversary has black box access to P, P^{-1}, and to an encryption and decryption oracle) [7]. The only assumption is that P is chosen uniformly at random from the set of all permutations of $\{0,1\}^n$ before it is published (obviously, the EM cipher is not secure with *any* choice of P). This bound is tight: Daemen [3] showed a chosen-plaintext key-recovery attack after $2^{n/2}$ evaluations of P and the encryption oracle.

Bogdanov et al. [2] generalized the EM construction to an r-rounds iterated EM cipher (with $r = 1$, it is the original EM cipher). This cipher uses r publicly known random permutations of $\{0,1\}^n$, $P_1, P_2, \ldots, P_r$, and $r+1$ secret keys $k_1, k_2, \ldots, k_r, k_{r+1} \in \{0,1\}^n$. Encryption of an n-bit plaintext m is carried out by the transformation

$$EM_{P_1,P_2,\ldots,P_r;k_1,k_2,\ldots,k_{r+1}}(m) = P_r(\ldots P_2(P_1(m \oplus k_1) \oplus k_2) \ldots \oplus k_r) \oplus k_{r+1}$$

(decryption is obvious). They showed that the r-rounds EM cipher ($r \geq 2$) is secure in the following sense: an adversary who sees no more than $2^{2n/3}$ chosen plaintext-ciphertext pairs cannot distinguish the encryption oracle from a random permutation of $\{0,1\}^n$.

As a practical example, Bogdanov et al. defined the 128-bit block cipher AES^2, which is an instantiation of the 2-rounds EM cipher where the two public permutations are AES (encryption) with two publicly known "arbitrary" keys (they chose the binary digits of the constant π). The complexity of the best (meet-in-the-middle) attack they showed used $2^{129.6}$ cipher revaluations. Consequently, they conjectured that AES^2 offers 128-bit security.

Dinur et al. [6] designed a chosen-plaintext key-recovery attack on a single key variant (i.e., the choice of $k_1 = k_2 = k_3 = k_4$) of the 3-rounds and also 2-rounds EM cipher, in time $O\left(\frac{\log n}{n} 2^n\right)$. This attack is based on the observation that for a "typical" permutation P of $\{0,1\}^n$, the distribution of $P(x) \oplus x$ over uniformly chosen $x \in \{0,1\}^n$ is not uniform. Their paper attributes this observation to Nikolić et al. [10], who demonstrated a chosen-plaintext key-recovery attack on the 2-rounds EM cipher with complexity slightly lower than exhaustive key search. In addition, [6] used the same observation to describe an attack on AES^2 (with three different keys) that is ~ 7 times faster than the best known meet-in-the-middle attack of Bogdanov et al. [2], therefore invalidating their conjecture on the security of AES^2. Although these attacks are not really practical, and not far from an exhaustive search, they demonstrates that the observation of Nikolić et al. can open a door to some potential threats.

In this paper we define a new variation of the EM cipher, which is immune against attacks based on any non-uniform properties of $P(x) \oplus x$. To this end, we first introduce the notion of a "balanced permutation", which is a permutation P of $\{0,1\}^n$ where $P(x) \oplus x$ is also a permutation. We show how to generate a large family of balanced permutations by using any permutation of $\{0,1\}^{n/2}$ in a Feistel construction. These can be used for constructing a 2-rounds balanced permutations EM block cipher with block size of $2n$ bits, using two public permutations of $\{0,1\}^n$. We prove that this block cipher is secure in the following sense: an adversary needs at least $\Omega(2^{n/2})$ chosen plaintext-ciphertext pairs before he can distinguish the block cipher from a random permutation of $\{0,1\}^{2n}$. Finally, we discuss a practical use of our construction, to define a 256-bit block cipher that is based on AES, and demonstrate its performance.

2 Balanced permutations and balanced permutations EM ciphers

2.1 Balanced permutations

Definition 1 (Balanced permutation). *Let σ be a permutation of $\{0,1\}^n$. Define the function $\widetilde{\sigma} : \{0,1\}^n \to \{0,1\}^n$ by $\widetilde{\sigma}(\omega) = \omega \oplus \sigma(\omega)$, for any $\omega \in \{0,1\}^n$. We say that σ is a balanced permutation if $\widetilde{\sigma}$ is also a permutation.*

Example 1. Let $A \in M_{n\times n}(\mathbb{Z}_2)$ be a matrix such that both A and $I + A$ are invertible. Define $\pi_A : \mathbb{Z}_2^n \to \mathbb{Z}_2^n$ by $\pi_A(x) = Ax$. Then π_A is a balance permutation of $\{0,1\}^n$. One such matrix is defined by $A_{i,i} = A_{i,i+1} = 1$ for $i = 1, 2, \ldots, n-1$, $A_{n,1} = 1$ and $A_{i,j} = 0$ for all other $1 \leq i, j \leq n$.

Example 2. Let a be an element of $GF(2^n)$ such that $a \neq 0, 1$. Identify $GF(2^n)$ with $\{0,1\}^n$, so field addition corresponds to bitwise XOR. The field's multiplication is denoted by $\times$. The function $x \to a \times x$ is a balanced permutation of $\{0,1\}^n$. Note that this example is actually a special case of the previous one.

The balanced permutations provided by the above examples are a small family of permutations, and moreover are all linear. We now give a recipe for generating a large family of balanced permutations, by employing the Feistel construction that turns any function $f : \{0,1\}^n \to \{0,1\}^n$ to a permutation of $\{0,1\}^{2n}$.

Let us use the following notations. For a string $\omega \in \{0,1\}^{2n}$, denote the string of its first n bits by $\omega_L \in \{0,1\}^n$, and the string of its last n bits by $\omega_R \in \{0,1\}^n$. Denote the concatenation of two strings $\omega_1, \omega_2 \in \{0,1\}^n$ (in this order) by $\omega_1 * \omega_2 \in \{0,1\}^{2n}$. We have the following identities:

$$(\omega_1 * \omega_2)_L = \omega_1, \quad (\omega_1 * \omega_2)_R = \omega_2, \quad \omega_L * \omega_R = \omega \tag{1}$$

Definition 2 (2 Feistel rounds permutation). *Let $f : \{0,1\}^n \to \{0,1\}^n$ be any function, and let $\pi_f : \{0,1\}^{2n} \to \{0,1\}^{2n}$ be the function*

$$\pi_f(\omega) := \omega_R * (\omega_L \oplus f(\omega_R)) . \tag{2}$$

The permutation $\pi_f^2 := \pi_f \circ \pi_f$ of $\{0,1\}^{2n}$ is called a 2 Feistel rounds permutation.

Clearly, π_f (the well known Feistel round) is a permutation of $\{0,1\}^{2n}$, and thus the composition π_f^2 (i.e., two Feistel rounds) is also a permutation of $\{0,1\}^{2n}$. We prove the following property for the special case where the function f is a permutation.

Proposition 1. *Let f be a permutation of $\{0,1\}^n$. Then, the 2 Feistel rounds permutation π_f^2 is a balanced permutation of $\{0,1\}^{2n}$. We call it a 2 Feistel rounds balanced permutation.*

Proof. First observe that

$$\pi_f^2(\omega) = \pi_f(\pi_f(\omega)) = \pi_f\left(\omega_R * (\omega_L \oplus f(\omega_R))\right) = \\ = (\omega_L \oplus f(\omega_R)) * (\omega_R \oplus f\left(\omega_L \oplus f(\omega_R)\right)). \quad (3)$$

Therefore,

$$\widetilde{\pi_f^2}(\omega) = f(\omega_R) * f\left(\omega_L \oplus f(\omega_R)\right).$$

Assume that $x, y \in \{0,1\}^{2n}$ such that $\widetilde{\pi_f^2}(x) = \widetilde{\pi_f^2}(y)$, i.e.,

$$f(x_R) * f\left(x_L \oplus f(x_R)\right) = f(y_R) * f\left(y_L \oplus f(y_R)\right)$$

Then, $f(x_R) = f(y_R)$ and $f\left(x_L \oplus f(x_R)\right) = f\left(y_L \oplus f(y_R)\right)$. Since (by assumption) f is one-to-one, $x_R = y_R$ and $x_L \oplus f(x_R) = y_L \oplus f(y_R)$, it follows that $x_L = (x_L \oplus f(x_R)) \oplus f(x_R) = (y_L \oplus f(y_R)) \oplus f(y_R) = y_L$. We established that $\widetilde{\pi_f^2}(x) = \widetilde{\pi_f^2}(y)$ implies $x = x_L * x_R = y_L * y_R = y$ which concludes the proof. □

Figure 1 shows an illustration of 2 Feistel rounds (balanced) permutation.

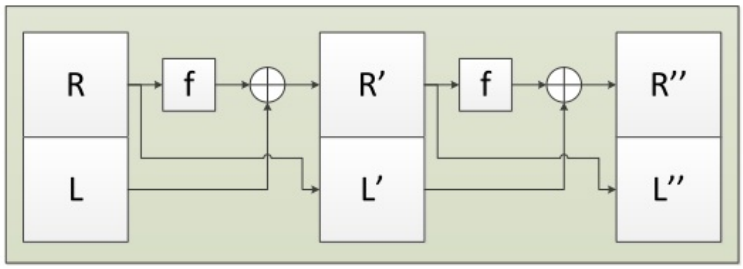

Fig. 1. The figure shows a function from $\{0,1\}^{2n}$ to $\{0,1\}^{2n}$, based on two Feistel rounds with a function $f : \{0,1\}^n \to \{0,1\}^n$. For any function f, this construction is a permutation of $\{0,1\}^{2n}$, denoted π_{f^2}. We call it a "2 Feistel rounds permutation". Proposition 1, shows that if f itself is a *permutation* of $\{0,1\}^n$, then π_{f^2} is a balanced permutation of $\{0,1\}^{2n}$. We call it a "2 Feistel rounds balanced permutation".

2.2 Balanced permutations EM ciphers

Definition 3 (r-rounds balanced permutations EM ciphers (BPEM)). *Let $n \geq 1$ and $r \geq 1$ be integers. Let $k_1, k_2, \ldots, k_{r+1}$ be $r+1$ string in $\{0,1\}^{2n}$. Let f_1,f_2,..., f_r be r permutations of $\{0,1\}^n$. Their associated 2 Feistel rounds balanced permutations (of $\{0,1\}^{2n}$) by $\pi_{f_1}^2$, $\pi_{f_2}^2$,...,$\pi_{f_r}^2$, respectively.*
The r-rounds balanced permutations EM (BPEM) block cipher is $EM_{\pi_{f_1}^2,\pi_{f_2}^2,\ldots,\pi_{f_r}^2,k_1,k_2,\ldots,k_{r+1}}$ It encrypts $2n$-bit blocks with an r-rounds EM cipher with the keys $k_1, k_2, \ldots, k_{r+1}$, where the r permutations $P_1, P_2, \ldots, P_r$ (of

$\{0,1\}^{2n}$) *are set to be* $\pi^2_{f_1}, \pi^2_{f_2}, \ldots, \pi^2_{f_r}$, *respectively.*
The use of the r*-rounds BPEM cipher for encryption (and decryption) starts with an initialization step, where the permutations* $f_1, f_2, \ldots, f_r$ *are selected uniformly and independently, at random from the set of permutations of* $\{0,1\}^n$. *After they are selected, they can be made public. Subsequently, per session/message, the secret keys* $k_1, k_2, \ldots, k_{r+1}$ *are selected uniformly and independently, at random, from* $\{0,1\}^{2n}$.

Figure 2 illustrates a 2-rounds BPEM cipher, which is the focus of this paper.

Remark 1. The 1-round BPEM cipher does not satisfy the requirement for a "random permutation selection" as in the definition of the EM scheme. Not surprisingly, it is easy to see that the 1-round BPEM does not preserve plaintext confidentiality. For any plaintexts $m \in \{0,1\}^{2n}$, we have

$$\left(\pi^2_f(m \oplus k_1)\right)_L = (m_L \oplus (k_1)_L) \oplus f(m_R \oplus (k_1)_R)$$

Therefore,

$$\left(EM_{\pi^2_f;k_1,k_2}(m)\right)_L = \left(\pi^2_f(m \oplus k_1)\right)_L \oplus (k_2)_L =$$
$$= m_L \oplus (k_1)_L \oplus (k_2)_L \oplus f(m_R \oplus (k_1)_R).$$

It follows that if, e.g., $(m_1)_R = (m_2)_R$ then

$$\left(EM_{\pi^2_f;k_1,k_2}(m_1) \oplus EM_{\pi^2_f;k_1,k_2}(m_2)\right)_L = (m_1 \oplus m_2)_L$$

which means that the ciphertexts leak out information on m_1, m_2. This implies that the r-rounds BPEM cipher must be used with $r \geq 2$ to have any hope for achieving security.

Remark 2. The r-rounds BPEM cipher is not necessarily secure with *any* choice of balanced permutations as $\pi^2_{f_1}, \pi^2_{f_2}, \ldots, \pi^2_{f_r}$, even if $r \geq 2$. For example, the cipher can be easily broken when using the linear balanced permutations shown in Examples 1 and 2.

Remark 3. By construction, $EM_{\pi^2_{f_1},\pi^2_{f_2},\ldots,\pi^2_{f_r},k_1,k_2,\ldots,k_{r+1}}$ $(r \geq 2)$ is immune against any attack that tries to leverage the non-uniformity of $P(x) \oplus x$ (including [10] and [6])). Obviously, this does not guarantee it is secure (as indicated in Remark 2).

In the next section, we prove that the 2-round EMBP cipher is indistinguishable from a random permutation.

3 Indistinguishability analysis for the 2-rounds BPEM cipher

Consider an oracle that chooses uniformly and independently, at random, permutations f_1, f_2 of $\{0,1\}^n$, a permutation P of $\{0,1\}^{2n}$ and keys k_1, k_2, k_3

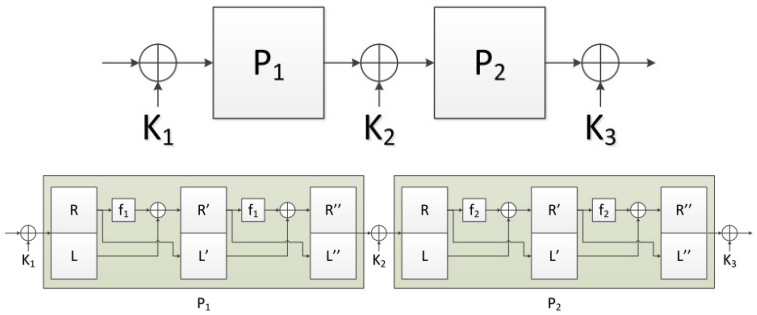

Fig. 2. The 2-rounds balanced permutations EM (BPEM) cipher operates on blocks of size $2n$ bits. The permutations P_1 and P_2 are balanced permutations of $\{0,1\}^{2n}$, defined as 2 Feistel rounds permutations. f_1 and f_2 are two (public) permutations of $\{0,1\}^n$. Each one of k_1, k_2, k_2 is a $2n$-bit secret key. See explanation in the text.

in $\{0,1\}^{2n}$. Then it selects a permutation F of $\{0,1\}^{2n}$, which is either P or $EM_{\pi^2_{f_1},\pi^2_{f_2};k_1,k_2,k_3}$.

An adversary selects a "querying and guessing" algorithm. He first uses it to submit to the oracle an adaptive series of queries of types f_1, f_2 and F. Queries of type f_i are of the form $f_i(x) =?$ or $f_i^{-1}(x) =?$ for $x \in \{0,1\}^n$ and queries of type F are of the form $F(w) =?$, $F^{-1}(w) =?$ for $w \in \{0,1\}^{2n}$. After collecting the responses, the adversary uses his algorithm to guess whether F is $EM_{\pi^2_{f_1},\pi^2_{f_2};k_1,k_2,k_3}$ or P. The quality of such an algorithm (in the cryptographic context) is the ability to distinguish between the two cases (rather than to successfully guess 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 F to be $EM_{\pi^2_{f_1},\pi^2_{f_2};k_1,k_2,k_3}$, and the probability that the algorithm outputs the same answer, given that the oracle chose F to be the random permutation. This is called the "advantage" of the algorithm. We are interested in bounding $Adv^{EM}_{f_1,f_2;k_1,k_2,k_3}(q_1,q_2,q_F)$, which is the maximal advantage of the adversary, over all possible algorithms, as a function of the budget q_1 of queries of type f_1, the budget q_2 of queries of type f_2 and the budget q_F of queries of type F. The total budget of queries is thus $q := q_1 + q_2 + q_F$. We show

Theorem 1. *If* $2(q+q_F)q_F < 2^n$ *then*

$$Adv^{EM}_{f_1,f_2;k_1,k_2,k_3}(q_1,q_2,q_F) \leq \frac{(6q+5q_F)q_F}{2^n - 2(q_F+q)q_F}.$$

and from the other direction, we also show

Proposition 2. *If* $q_F \leq 2 \cdot 2^n$ *then*

$$Adv^{EM}_{f_1,f_2;k_1,k_2,k_3}(q_1,q_2,q_F) \geq \frac{\frac{1}{8}q_F^2}{2^{2n}}.$$

For the proof of Theorem 1 we consider an algorithm that achieves the largest advantage. We may assume that:

1. The adversary never repeats a query that was already made, and does not make '"opposite" queries (i.e., $\pi^{-1}(x) =?$ after already receiving x as a reply to a query $\pi(\cdot) =?$ where π is one of $f_1, f_1^{-1}, f_2, f_2^{-1}, F, F^{-1}$).
2. By adding superfluous queries whose answers the algorithm will ignore, we may assume the adversary makes exactly q_1 queries of type f_1, exactly q_2 queries of type f_2 and exactly q_F queries of type F.
3. The adversary's algorithm, and the choice of the queries in particular, are completely deterministic. This means that the first query the adversary makes is fixed, and any subsequent query he makes depends only on the answers he got for the previous queries. Therefore, if the oracle uses $F = EM_{\pi^2_{f_1},\pi^2_{f_2};k_1,k_2,k_3}$, for some choice of f_1, f_2, k_1, k_2, k_3, then all queries and their answers are completely determined by f_1, f_2, k_1, k_2, k_3. We denote this sequence of queries and answers by $qa(f_1, f_2; k_1, k_2, k_3)$.

Let QA be the set of all possible sequences of query-answer pairs that can be obtained through the interaction between the adversary and the oracle. For any $\bar{s} \in QA$, we denote by $pr_{perm}(\bar{s})$ the probability that $\bar{s}$ will be the sequence of pairs obtained by the adversary-oracle interaction when the oracle uses a random permutation as F. Similarly, $\Pr_{EM}(\bar{s})$ is the probability that $\bar{s}$ will be the sequence of pairs obtained by the adversary-oracle interaction when the oracle uses a balanced EM block cipher as F.

Let $E \subseteq QA$ be the set of sequences of query-answer pairs for which the adversary will guess that F is a balanced EM block cipher. The advantage of the algorithm is $|\Pr_{EM}(E) - \Pr_{perm}(E)|$.

Let us say that the pair $< x, z >$ is established by a query if either the query is $\pi(x) =?$ (where π is either F, f_1 or f_2) and its answer is z, or if the query is $\pi^{-1}(z) =?$ (where, again, π is either F, f_1 or f_2) and its answer is x.

For an $\bar{s} \in QA$, let

$$(u_1^1(\bar{s}), v_1^1(\bar{s})), (u_2^1(\bar{s}), v_2^1(\bar{s})), \ldots, (u_{q_1}^1(\bar{s}), v_{q_1}^1(\bar{s}))$$

be the pairs established by queries of type f_1 in $\bar{s}$, let

$$(u_1^2(\bar{s}), v_1^2(\bar{s})), (u_2^2(\bar{s}), v_2^2(\bar{s})), \ldots, (u_{q_2}^2(\bar{s}), v_{q_2}^2(\bar{s}))$$

be the pairs established by queries of type f_2 in $\bar{s}$. Let

$$(x_1(\bar{s}), z_1(\bar{s})), (x_2(\bar{s}), z_2(\bar{s})), \ldots, (x_{q_F}(\bar{s}), z_{q_F}(\bar{s}))$$

be the pairs established by queries of type F in $\bar{s}$. Let $q_{F,1}(\bar{s}) := |\{(x_i(\bar{s}))_R\}_{i=1}^{q_F}|$, $q_{F,2}(\bar{s}) := |\{(z_i(\bar{s}))_L\}_{i=1}^{q_F}|$.

Let S_{2^n} be the set of permutations of $\{0,1\}^n$. For any $\bar{s} \in QA$, $\bar{k} = (k_1, k_2, k_3) \in \left(\{0,1\}^{2n}\right)^3$ and $\bar{y} = (y_1, y_2, \ldots, y_{q_F}) \in \left(\{0.1\}^{2n}\right)^{q_F}$ let

$$em(\bar{s}, \bar{k}, \bar{y}) := \{(f_1, f_2) \in (S_{2^n})^2 \mid qa(f_1, f_2; k_1, k_2, k_3) = \bar{s}\,,\ \forall 1 \le i \le q_F :$$
$$f_1((x_i(\bar{s}) \oplus k_1)_R) = (y_i \oplus x_i(\bar{s}) \oplus k_1)_L,\ f_2\left((z_i(\bar{s}) \oplus k_3)_L\right) = (y_i \oplus z_i(\bar{s}) \oplus k_2 \oplus k_3)_R\}.$$

It is easy to see that if $em(\bar{s},\bar{k},\bar{y}) \neq \emptyset$ then for any $1 \leq i < j \leq q_F$,

$$(x_i(\bar{s}))_R = (x_j(\bar{s}))_R \leftrightarrow (y_i \oplus x_i(\bar{s}))_L = (y_j \oplus x_j(\bar{s}))_L,$$

and

$$(y_i \oplus z_i(\bar{s}))_R = (y_j \oplus z_j(\bar{s}))_R \leftrightarrow (z_i(\bar{s}))_L = (z_j(\bar{s}))_L.$$

Let $Y(\bar{s})$ be the set of all $\bar{y}$'s in $\left(\{0,1\}^{2n}\right)^{q_F}$ satisfying those conditions.

For any $S \subseteq QA$, let

$$\widehat{S} := \{(\bar{s},\bar{k},\bar{y}) \mid \bar{s} \in S\,,\, \bar{k} \in \left(\{0,1\}^{2n}\right)^3, \bar{y} \in Y(\bar{s})\}.$$

For any $(\bar{s},\bar{k},\bar{y}) \in \widehat{QA}$ let

$$\alpha(\bar{s},\bar{k},\bar{y}) := \frac{|em(\bar{s},\bar{k},\bar{y})|}{\left((2^n)!\right)^2 \left(2^{2n}\right)^3}.$$

For any $\bar{s} \in QA$ let

$$\beta(\bar{s}) := \frac{\mathrm{Pr}_{perm}(\bar{s})}{\left(2^{2n}\right)^3 |Y(\bar{s})|}.$$

Lemma 1. *For any $S \subseteq QA$*

$$\mathrm{Pr}_{EM}(S) - \mathrm{Pr}_{perm}(S) = \sum_{(\bar{s},\bar{k},\bar{y}) \in \widehat{S}} \left(\alpha(\bar{s},\bar{k},\bar{y}) - \beta(\bar{s})\right).$$

Proof. For any $\bar{s} \in QA$

$$\mathrm{Pr}_{EM}(\bar{s}) = \sum_{\substack{\bar{k} \in \left(\{0,1\}^{2n}\right)^3 \\ \bar{y} \in Y(\bar{s})}} \alpha(\bar{s},\bar{k},\bar{y}),$$

therefore

$$\mathrm{Pr}_{EM}(S) - \mathrm{Pr}_{perm}(S) = \sum_{\bar{s} \in S} \left(\mathrm{Pr}_{EM}(\bar{s}) - \mathrm{Pr}_{perm}(\bar{s})\right) =$$

$$= \sum_{\bar{s} \in S} \sum_{\substack{\bar{k} \in \left(\{0,1\}^{2n}\right)^3 \\ \bar{y} \in Y(\bar{s})}} \left(\alpha(\bar{s},\bar{k},\bar{y}) - \beta(\bar{s})\right) = \sum_{(\bar{s},\bar{k},\bar{y}) \in \widehat{S}} \left(\alpha(\bar{s},\bar{k},\bar{y}) - \beta(\bar{s})\right).$$

□

In Subsection 3.1 we define a set $D_* \subseteq \widehat{QA}$ of "unfortunate instances", and prove the following two Lemmas.

Lemma 2.

$$\sum_{(\bar{s},\bar{k},\bar{y}) \in D_*} \beta(\bar{s}) \leq \frac{(4q + 3q_F)q_F}{2^n - q_F}$$

Lemma 3. *Suppose* $2(q+q_F)q_F < 2^n$. *For any* $(\bar{s},\bar{k},\bar{y}) \in \widehat{QA} - D_*$,

$$\left|\frac{\alpha(\bar{s},\bar{k},\bar{y})}{\beta(\bar{s})} - 1\right| < \frac{2(q_F+q)q_F}{2^n - 2(q_F+q)q_F}$$

The proof of Theorem 1 follows via a standard argument which we describe in Subsection 3.2.

We comment that up to some obvious changes, the same arguments show that the same bound holds for the single key variation of the 2-rounds BPEM cipher. To put it formally, we consider an oracle that chooses uniformly and independently, at random, permutations f_1, f_2 of $\{0,1\}^n$, a permutation P of $\{0,1\}^{2n}$ and a *single* key $k \in \{0,1\}^{2n}$. Then, it selects a permutation F of $\{0,1\}^{2n}$, which is either P or $EM_{\pi^2_{f_1},\pi^2_{f_2};k,k,k}$. As before, the adversary is expected to guess whether F is $EM_{\pi^2_{f_1},\pi^2_{f_2};k,k,k}$ or P, after collecting the responses to q queries, where q_F of them are of type F. Let $Adv^{EM}_{f_1,f_2;k,k,k}(q,q_F)$ be the maximal advantage of the adversary, over all possible algorithms.

We have then

Theorem 2. *If* $2(q+q_F)q_F < 2^n$ *then*

$$Adv^{EM}_{f_1,f_2;k,k,k}(q,q_F) \leq \frac{(6q+5q_F)q_F}{2^n - 2(q_F+q)q_F}.$$

We also comment that similar arguments yield bounds for the following variations of the 2-rounds BPEM cipher: single permutation, and both single key and single permutation (the formal description is analogous to the one above).

Theorem 3. *If* $4(q+q_F)q_F < 2^n$ *then*

$$Adv^{EM}_{f,f;k_1,k_2,k_3}(q,q_F) \leq \frac{(12q+11q_F)q_F}{2^n - 4(q_F+q)q_F}.$$

Theorem 4. *If* $4(q+q_F)q_F < 2^n$ *then*

$$Adv^{EM}_{f,f;k,k,k}(q,q_F) \leq \frac{(12q+11q_F)q_F}{2^n - 4(q_F+q)q_F}.$$

3.1 Definition of the set D_* and proofs of its properties

The set D_* will be the union of all the sets defined below.

Definition 4. *For* $1 \leq i \leq q_1$, $1 \leq j \leq q_F$ *let*

$$D^{u^1,R}_{i,j} := \{(\bar{s},\bar{k},\bar{y}) \in \widehat{QA} \mid u^1_i(\bar{s}) = (x_j(\bar{s}) \oplus k_1)_R\},$$
$$D^{u^1,L}_{i,j} := \{(\bar{s},\bar{k},\bar{y}) \in \widehat{QA} \mid u^1_i(\bar{s}) = (y_j)_L\},$$
$$D^{v^1,L}_{i,j} := \{(\bar{s},\bar{k},\bar{y}) \in \widehat{QA} \mid v^1_i(\bar{s}) = (y_j \oplus x_j(\bar{s}) \oplus k_1)_L\},$$
$$D^{v^1,R}_{i,j} := \{(\bar{s},\bar{k},\bar{y}) \in \widehat{QA} \mid v^1_i(\bar{s}) = (y_j \oplus x_j(\bar{s}) \oplus k_1)_R\},$$

for $1 \le i \le q_2$, $1 \le j \le q_F$ *let*

$$D^{u^2,L}_{i,j} := \{(\bar{s},\bar{k},\bar{y}) \in \widehat{QA} \mid u^2_i(\bar{s}) = (z_j(\bar{s}) \oplus k_3)_L\},$$
$$D^{u^2,R}_{i,j} := \{(\bar{s},\bar{k},\bar{y}) \in \widehat{QA} \mid u^2_i(\bar{s}) = (y_j \oplus k_2)_R\},$$
$$D^{v^2,L}_{i,j} := \{(\bar{s},\bar{k},\bar{y}) \in \widehat{QA} \mid v^2_i(\bar{s}) = (y_j \oplus z_j(\bar{s}) \oplus k_2 \oplus k_3)_L\},$$
$$D^{v^2,R}_{i,j} := \{(\bar{s},\bar{k},\bar{y}) \in \widehat{QA} \mid v^2_i(\bar{s}) = (y_j \oplus z_j(\bar{s}) \oplus k_2 \oplus k_3)_R\},$$

for $1 \le i,j \le q_F$ *let*

$$D^{y_L,x_R}_{i,j} := \{(\bar{s},\bar{k},\bar{y}) \in \widehat{QA} \mid (y_i)_L = (x_j(\bar{s}) \oplus k_1)_R\},$$
$$D^{y_R,z_L}_{i,j} := \{(\bar{s},\bar{k},\bar{y}) \in \widehat{QA} \mid (y_i \oplus k_2)_R = (z_j(\bar{s}) \oplus k_3)_L\},$$
$$D^{y,x}_{i,j} := \{(\bar{s},\bar{k},\bar{y}) \in \widehat{QA} \mid (y_i \oplus x_i(\bar{s}))_L = (y_j \oplus x_j(\bar{s}))_R\},$$
$$D^{y,z}_{i,j} := \{(\bar{s},\bar{k},\bar{y}) \in \widehat{QA} \mid (y_i \oplus z_i(\bar{s}))_L = (y_j \oplus z_j(\bar{s}))_R\},$$

and finally, for $1 \le i < j \le q_F$ *let*

$$D^{y_L}_{i,j} := \{(\bar{s},\bar{k},\bar{y}) \in \widehat{QA} \mid (y_i)_L = (y_j)_L\},$$
$$D^{y_R}_{i,j} := \{(\bar{s},\bar{k},\bar{y}) \in \widehat{QA} \mid (y_i)_R = (y_j)_R\},$$
$$D^{y_R,x_R,\neq}_{i,j} := \{(\bar{s},\bar{k},\bar{y}) \in \widehat{QA} \mid (y_i \oplus x_i(\bar{s}))_R = (y_j \oplus x_j(\bar{s}))_R\,,\,(y_i \oplus z_i(\bar{s}))_R \neq (y_j \oplus z_j(\bar{s}))_R\},$$
$$D^{y_L,z_L,\neq}_{i,j} := \{(\bar{s},\bar{k},\bar{y}) \in \widehat{QA} \mid (y_i \oplus z_i(\bar{s}))_L = (y_j \oplus z_j(\bar{s}))_L\,,\,(y_i \oplus x_i(\bar{s}))_L \neq (y_j \oplus x_j(\bar{s}))_L\},$$
$$D^{y_R,x_R,=}_{i,j} := \{(\bar{s},\bar{k},\bar{y}) \in \widehat{QA} \mid (y_i \oplus x_i(\bar{s}))_R = (y_j \oplus x_j(\bar{s}))_R\,,\,(y_i \oplus z_i(\bar{s}))_R = (y_j \oplus z_j(\bar{s}))_R\},$$
$$D^{y_L,z_L,=}_{i,j} := \{(\bar{s},\bar{k},\bar{y}) \in \widehat{QA} \mid (y_i \oplus z_i(\bar{s}))_L = (y_j \oplus z_j(\bar{s}))_L\,,\,(y_i \oplus x_i(\bar{s}))_L = (y_j \oplus x_j(\bar{s}))_L\}.$$

The number of sets defined in Definition 4 is

$$4q_1q_F + 4q_2q_F + 4q_F^2 + 6\binom{q_F}{2} < (4q + 3q_F)q_F,$$

Therefore, Lemma 2 follows directly from the following lemma.

Lemma 4. *Let D be any of the sets defined in Definition 4. Then*

$$\sum_{(\bar{s},\bar{k},\bar{y}) \in D} \beta(\bar{s}) \le \frac{1}{2^n - q_F} \tag{4}$$

Proof. To show (4), it is enough to show that either

$$\sum_{\substack{\bar{s} \in QA \\ \exists \bar{k} \in \left(\{0,1\}^{2n}\right)^3, \bar{y} \in Y(\bar{s}) : (\bar{s},\bar{k},\bar{y}) \in D}} \Pr_{perm}(\bar{s}) \le \frac{1}{2^n - q_F},$$

or that for any $\bar{s} \in QA$,

$$\sum_{\substack{\bar{y} \in Y(\bar{s}) \\ \exists \bar{k} \in \left(\{0,1\}^{2n}\right)^3 : (\bar{s}, \bar{k}, \bar{y}) \in D}} \frac{1}{|Y(\bar{s})|} \leq \frac{1}{2^n - q_F},$$

or that for any $\bar{s} \in QA$ and $\bar{y} \in Y(\bar{s})$,

$$\sum_{\substack{\bar{k} \in \left(\{0,1\}^{2n}\right)^3 \\ (\bar{s}, \bar{k}, \bar{y}) \in D}} \frac{1}{(2^{2n})^3} \leq \frac{1}{2^n - q_F}.$$

We now verify that for any set D, defined in Definition 4, at least one of the above holds.

1. If D is any of the sets $D_{i,j}^{u^1,R}, D_{i,j}^{v^1,L}, D_{i,j}^{v^1,R}, D_{i,j}^{u^2,L}, D_{i,j}^{u^2,R}, D_{i,j}^{y_L,x_R}$, then for any $\bar{s} \in QA$ and $\bar{y} \in Y(\bar{s})$,

$$\sum_{\substack{\bar{k} \in \left(\{0,1\}^{2n}\right)^3 \\ (\bar{s}, \bar{k}, \bar{y}) \in D}} \frac{1}{(2^{2n})^3} = \frac{1}{2^n}.$$

2. If D is any of the sets $D_{i,j}^{u^1,L}, D_{i,j}^{v^2,L}, D_{i,j}^{y_L,z_L,\neq}$ then for any $\bar{s} \in QA$,

$$\sum_{\substack{\bar{y} \in Y(\bar{s}) \\ \exists \bar{k} \in \left(\{0,1\}^{2n}\right)^3 : (\bar{s}, \bar{k}, \bar{y}) \in D}} \frac{1}{|Y(\bar{s})|} = \frac{1}{2^n - (q_{F,1}(\bar{s}) - 1)}.$$

   In the case that $D = D_{i,j}^{y_L,z_L,\neq}$ we rely on the observation that the condition $(y_i \oplus x_i(\bar{s}))_L \neq (y_j \oplus x_j(\bar{s}))_L$ ensures that the condition $(y_i \oplus z_i(\bar{s}))_L = (y_j \oplus z_j(\bar{s}))_L$ is independent of the conditions defining $Y(\bar{s})$.
3. If D is any of the sets $D_{i,j}^{v^2,R}, D_{i,j}^{y_R,z_L}, D_{i,j}^{y_R,x_R,\neq}$ then for any $\bar{s} \in QA$,

$$\sum_{\substack{\bar{y} \in Y(\bar{s}) \\ \exists \bar{k} \in \left(\{0,1\}^{2n}\right)^3 : (\bar{s}, \bar{k}, \bar{y}) \in D}} \frac{1}{|Y(\bar{s})|} = \frac{1}{2^n - (q_{F,2}(\bar{s}) - 1)}.$$

   In the case that $D = D_{i,j}^{y_R,x_R,\neq}$ we rely on the observation that the condition $(y_i \oplus z_i(\bar{s}))_R \neq (y_j \oplus z_j(\bar{s}))_R$ ensures that the condition $(y_i \oplus x_i(\bar{s}))_R = (y_j \oplus x_j(\bar{s}))_R$ is independent of the conditions defining $Y(\bar{s})$.
4. If D is any of the sets $D_{i,j}^{y,x}, D_{i,j}^{y,z}$ then for any $\bar{s} \in QA$,

$$\sum_{\substack{\bar{y} \in Y(\bar{s}) \\ \exists \bar{k} \in \left(\{0,1\}^{2n}\right)^3 : (\bar{s}, \bar{k}, \bar{y}) \in D}} \frac{1}{|Y(\bar{s})|} = \frac{1}{2^n}.$$

5. If D is any of the sets $D_{i,j}^{y_L}$ then for any $\bar{s} \in QA$,

$$\sum_{\substack{\bar{y} \in Y(\bar{s}) \\ \exists \bar{k} \in \left(\{0,1\}^{2n}\right)^3 : (\bar{s}, \bar{k}, \bar{y}) \in D}} \frac{1}{|Y(\bar{s})|} \leq \frac{1}{2^n - (q_{F,1}(\bar{s}) - 1)},$$

since the condition $(y_i)_L = (y_j)_L$ is either contradicting or independent of the conditions defining $Y(\bar{s})$

6. If D is any of the sets $D_{i,j}^{y_R}$ then for any $\bar{s} \in QA$,

$$\sum_{\substack{\bar{y} \in Y(\bar{s}) \\ \exists \bar{k} \in \left(\{0,1\}^{2n}\right)^3 : (\bar{s}, \bar{k}, \bar{y}) \in D}} \frac{1}{|Y(\bar{s})|} \leq \frac{1}{2^n - (q_{F,2}(\bar{s}) - 1)},$$

since the condition $(y_i)_R = (y_j)_R$ is either contradicting or independent of the conditions defining $Y(\bar{s})$

7. Finally, if D is any of the sets $D_{i,j}^{y_L, z_L, =}, D_{i,j}^{y_R, x_R, =}$ then since

$$D_{i,j}^{y_L, z_L, =} \subseteq \{(\bar{s}, \bar{k}, \bar{y}) \in \widehat{QA} \mid (x_i(\bar{s}) \oplus z_i(\bar{s}))_L = (x_j(\bar{s}) \oplus z_j(\bar{s}))_L\},$$
$$D_{i,j}^{y_R, x_R, =} \subseteq \{(\bar{s}, \bar{k}, \bar{y}) \in \widehat{QA} \mid (x_i(\bar{s}) \oplus z_i(\bar{s}))_R = (x_j(\bar{s}) \oplus z_j(\bar{s}))_R\},$$

we get that

$$\sum_{\substack{\bar{s} \in QA \\ \exists \bar{k} \in \left(\{0,1\}^{2n}\right)^3, \bar{y} \in Y(\bar{s}) : (\bar{s}, \bar{k}, \bar{y}) \in D}} \Pr_{perm}(\bar{s}) \leq \frac{2^n}{2^{2n} - 1}.$$

□

We now address the proof of Lemma 3. We use the following computational lemma.

Lemma 5. *For any integers $r, N \geq 1$ and $m \geq 0$,*

$$1 \geq \prod_{i=1}^{r} \left(1 - \frac{i + m - 1}{N}\right) > 1 - \frac{\frac{1}{2}(r + 2m)r}{N}.$$

Proof.

$$1 \geq \prod_{i=1}^{r} \left(1 - \frac{i + m - 1}{N}\right) = \sqrt{\prod_{i=1}^{r} \left(1 - \frac{i + m - 1}{N}\right) \left(1 - \frac{r - i + m}{N}\right)} \geq$$
$$\geq \left(1 - \frac{r + 2m - 1}{N}\right)^{\frac{r}{2}} \geq 1 - \frac{\frac{1}{2}(r + 2m - 1)r}{N} > 1 - \frac{\frac{1}{2}(r + 2m)r}{N}.$$

□

Proof of Lemma 3. A pair $(f_1, f_2) \in (S_{2^n})^2$ belongs to $em(\bar{s}, \bar{k}, \bar{y})$ if and only if f_1 satisfies:

- $f_1(u_i^1(\bar{s})) = v_i^1(\bar{s})$ for $1 \le i \le q_1$,
- $f_1((x_i(\bar{s}) \oplus k_1)_R) = (y_i \oplus x_i(\bar{s}) \oplus k_1)_L$ for $1 \le i \le q_F$,
- $f_1((y_i)_L) = (y_i \oplus x_i(\bar{s}) \oplus k_1)_R$ for $1 \le i \le q_F$,

and f_2 satisfies:

- $f_2(u_i^2(\bar{s})) = v_i^2(\bar{s})$ for $1 \le i \le q_2$,
- $f_2((y_i \oplus k_2)_R) = (y_i \oplus z_i(\bar{s}) \oplus k_2 \oplus k_3)_L$ for $1 \le i \le q_F$,
- $f_2((z_i(\bar{s}) \oplus k_3)_L) = (y_i \oplus z_i(\bar{s}) \oplus k_2 \oplus k_3)_R$ for $1 \le i \le q_F$.

The definitions of D_* and $Y(\bar{s})$ guarantees that for $(\bar{s}, \bar{k}, \bar{y}) \in \widehat{QA} - D_*$, these are exactly $q_1 + q_{F,1}(\bar{s}) + q_F$ non-contradicting requirements for f_1 (some of the requirements of the form $f_1((x_i(\bar{s}) \oplus k_1)_R) = (y_i \oplus x_i(\bar{s}) \oplus k_1)_L$ coincide), and exactly $q_2 + q_{F,2}(\bar{s}) + q_F$ non-contradicting requirements for f_2 (some of the requirements of the form $f_2((z_i(\bar{s}) \oplus k_3)_L) = (y_i \oplus z_i(\bar{s}) \oplus k_2 \oplus k_3)_R$ coincide). Hence

$$\left|em(\bar{s}, \bar{k}, \bar{y})\right| = (2^n - (q_1 + q_{F,1}(\bar{s}) + q_F))!\,(2^n - (q_2 + q_{F,2}(\bar{s}) + q_F))!.$$

Since $|Y(\bar{s})| = \prod_{i=1}^{q_{F,1}(\bar{s})} (2^n - (i-1)) \cdot \prod_{i=1}^{q_{F,2}(\bar{s})} (2^n - (i-1))$ and

$$\Pr_{perm}(\bar{s}) = \frac{1}{\prod_{i=1}^{q_1} (2^n - (i-1)) \cdot \prod_{i=1}^{q_2} (2^n - (i-1)) \cdot \prod_{i=1}^{q_F} (2^{2n} - (i-1))},$$

we get that

$$\frac{\alpha(\bar{s}, \bar{k}, \bar{y})}{\beta(\bar{s})} = \frac{\prod_{i=1}^{q_{F,1}(\bar{s})} \left(1 - \frac{i-1}{2^n}\right) \cdot \prod_{i=1}^{q_{F,2}(\bar{s})} \left(1 - \frac{i-1}{2^n}\right) \cdot \prod_{i=1}^{q_F} \left(1 - \frac{i-1}{2^{2n}}\right)}{\prod_{i=1}^{q_{F,1}(\bar{s})+q_F} \left(1 - \frac{i+q_1-1}{2^n}\right) \cdot \prod_{i=1}^{q_{F,2}(\bar{s})+q_F} \left(1 - \frac{i+q_2-1}{2^n}\right)}.$$

Hence by Lemma 5

$$\frac{\alpha(\bar{s}, \bar{k}, \bar{y})}{\beta(\bar{s})} \ge \prod_{i=1}^{q_{F,1}(\bar{s})} \left(1 - \frac{i-1}{2^n}\right) \cdot \prod_{i=1}^{q_{F,2}(\bar{s})} \left(1 - \frac{i-1}{2^n}\right) \cdot \prod_{i=1}^{q_F} \left(1 - \frac{i-1}{2^{2n}}\right) >$$

$$> \left(1 - \frac{\frac{1}{2} q_{F,1}(\bar{s})^2}{2^n}\right) \left(1 - \frac{\frac{1}{2} q_{F,2}(\bar{s})^2}{2^n}\right) \left(1 - \frac{\frac{1}{2} q_F^2}{2^{2n}}\right) > \left(1 - \frac{\frac{1}{2} q_F^2}{2^n}\right)^3 >$$

$$> 1 - \frac{\frac{3}{2} q_F^2}{2^n} > 1 - \frac{2(q_F + q) q_F}{2^n - 2(q_F + q) q_F}.$$

On the other hand, again by Lemma 5,

$$\prod_{i=1}^{q_{F,1}(\bar{s})+q_F} \left(1 - \frac{i + q_1 - 1}{2^n}\right) > 1 - \frac{\frac{1}{2}(q_{F,1}(\bar{s}) + q_F + 2q_1)(q_{F,1}(\bar{s}) + q_F)}{2^n} \ge$$

$$\ge 1 - \frac{2(q_F + q_1) q_F}{2^n}$$

and

$$\prod_{i=1}^{q_{F,2}(\bar{s})+q_F}\left(1-\frac{i+q_2-1}{2^n}\right)>1-\frac{\frac{1}{2}(q_{F,1}(\bar{s})+q_F+2q_2)(q_{F,1}(\bar{s})+q_F)}{2^n}\geq$$
$$\geq 1-\frac{2(q_F+q_2)q_F}{2^n},$$

therefore

$$\frac{\beta(\bar{s})}{\alpha(\bar{s},\bar{k},\bar{y})}\geq\prod_{i=1}^{q_{F,1}(\bar{s})+q_F}\left(1-\frac{i+q_1-1}{2^n}\right)\cdot\prod_{i=1}^{q_{F,2}(\bar{s})+q_F}\left(1-\frac{i+q_2-1}{2^n}\right)>$$
$$>\left(1-\frac{2(q_F+q_1)q_F}{2^n}\right)\left(1-\frac{2(q_F+q_2)q_F}{2^n}\right)\geq$$
$$\geq 1-\frac{2(2q_F+q_1+q_2)q_F}{2^n}=1-\frac{2(q_F+q)q_F}{2^n},$$

hence

$$\frac{\alpha(\bar{s},\bar{k},\bar{y})}{\beta(\bar{s})}<\frac{1}{1-\frac{2(q_F+q)q_F}{2^n}}=1+\frac{2(q_F+q)q_F}{2^n-2(q_F+q)q_F}$$

and the lemma follows.

□

3.2 Proof of Theorem 1

We first show a simple corollary of Lemma 3.

Corollary 1.

$$\sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{QA}-D_*}\left|\alpha(\bar{s},\bar{k},\bar{y})-\beta(\bar{s})\right|\leq\frac{2(q_F+q)q_F}{2^n-2(q_F+q)q_F}.$$

Proof. By Lemma 3,

$$\sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{QA}-D_*}\left|\alpha(\bar{s},\bar{k},\bar{y})-\beta(\bar{s})\right|=\sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{QA}-D_*}\beta(\bar{s})\left|\frac{\alpha(\bar{s},\bar{k},\bar{y})}{\beta(\bar{s})}-1\right|\leq$$
$$\leq\left(\sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{QA}-D_*}\beta(\bar{s})\right)\frac{2(q_F+q)q_F}{2^n-2(q_F+q)q_F}\leq\left(\sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{QA}}\beta(\bar{s})\right)\frac{2(q_F+q)q_F}{2^n-2(q_F+q)q_F}=$$
$$=\frac{2(q_F+q)q_F}{2^n-2(q_F+q)q_F}.$$

□

We will also use the following simple lemma

Lemma 6.

$$\sum_{(\bar{s},\bar{k},\bar{y})\in D_*} \alpha(\bar{s},\bar{k},\bar{y}) \le \sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{QA}-D_*} \left|\alpha(\bar{s},\bar{k},\bar{y}) - \beta(\bar{s})\right| + \sum_{(\bar{s},\bar{k},\bar{y})\in D_*} \beta(\bar{s}).$$

Proof. By Lemma 1,

$$\sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{QA}} \big(\alpha(\bar{s},\bar{k},\bar{y}) - \beta(\bar{s})\big) = \Pr_{EM}(QA) - \Pr_{perm}(QA) = 0.$$

Therefore, by Corollary 1,

$$\sum_{(\bar{s},\bar{k},\bar{y})\in D_*} \big(\alpha(\bar{s},\bar{k},\bar{y}) - \beta(\bar{s})\big) = -\sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{QA}-D_*} \big(\alpha(\bar{s},\bar{k},\bar{y}) - \beta(\bar{s})\big) \le$$
$$\le \sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{QA}-D_*} \left|\alpha(\bar{s},\bar{k},\bar{y}) - \beta(\bar{s})\right|.$$

hence

$$\sum_{(\bar{s},\bar{k},\bar{y})\in D_*} \alpha(\bar{s},\bar{k},\bar{y}) = \sum_{(\bar{s},\bar{k},\bar{y})\in D_*} \big(\alpha(\bar{s},\bar{k},\bar{y}) - \beta(\bar{s})\big) + \sum_{(\bar{s},\bar{k},\bar{y})\in D_*} \beta(\bar{s}) \le$$
$$\le \sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{QA}-D_*} \left|\alpha(\bar{s},\bar{k},\bar{y}) - \beta(\bar{s})\right| + \sum_{(\bar{s},\bar{k},\bar{y})\in D_*} \beta(\bar{s}).$$

□

Proof of Theorem 1. Recall that the advantage of the adversary's algorithm is $|\Pr_{EM}(E) - \Pr_{perm}(E)|$, where E is the set of sequences of queries and answers for which the adversary will guess that F is a balanced EM block cipher. For any $A \subseteq QA$, by Lemma 1,

$$|\Pr_{EM}(A) - \Pr_{perm}(A)| = \left|\sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{A}} \big(\alpha(\bar{s},\bar{k},\bar{y}) - \beta(\bar{s})\big)\right| =$$

$$= \left|\sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{A}-D_*} \big(\alpha(\bar{s},\bar{k},\bar{y}) - \beta(\bar{s})\big) + \sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{A}\cap D_*} \alpha(\bar{s},\bar{k},\bar{y}) - \sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{A}\cap D_*} \beta(\bar{s})\right| \le$$
$$\le \sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{A}-D_*} \left|\alpha(\bar{s},\bar{k},\bar{y}) - \beta(\bar{s})\right| + \sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{A}\cap D_*} \alpha(\bar{s},\bar{k},\bar{y}) + \sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{A}\cap D_*} \beta(\bar{s}).$$

In particular,

$$|\Pr_{EM}(E) - \Pr_{perm}(E)| \le$$
$$\le \sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{E}-D_*} \left|\alpha(\bar{s},\bar{k},\bar{y}) - \beta(\bar{s})\right| + \sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{E}\cap D_*} \alpha(\bar{s},\bar{k},\bar{y}) + \sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{E}\cap D_*} \beta(\bar{s}),$$

and for $E^c := QA - E$

$$|\Pr_{EM}(E^c) - \Pr_{perm}(E^c)| \leq$$
$$\leq \sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{E^c}-D_*} \left|\alpha(\bar{s},\bar{k},\bar{y}) - \beta(\bar{s})\right| + \sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{E^c}\cap D_*} \alpha(\bar{s},\bar{k},\bar{y}) + \sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{E^c}\cap D_*} \beta(\bar{s}),$$

hence, using Lemma 2, Corollary 1 and Lemma 6,

$$|\Pr_{EM}(E) - \Pr_{perm}(E)| + |\Pr_{EM}(E^c) - \Pr_{perm}(E^c)| \leq$$
$$\leq \sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{QA}-D_*} \left|\alpha(\bar{s},\bar{k},\bar{y}) - \beta(\bar{s})\right| + \sum_{(\bar{s},\bar{k},\bar{y})\in D_*} \alpha(\bar{s},\bar{k},\bar{y}) + \sum_{(\bar{s},\bar{k},\bar{y})\in D_*} \beta(\bar{s}) \leq$$
$$\leq 2 \sum_{(\bar{s},\bar{k},\bar{y})\in\widehat{QA}-D_*} \left|\alpha(\bar{s},\bar{k},\bar{y}) - \beta(\bar{s})\right| + 2 \sum_{(\bar{s},\bar{k},\bar{y})\in D_*} \beta(\bar{s}) \leq$$
$$\leq 2\frac{2(q_F+q)q_F}{2^n - 2(q_F+q)q_F} + 2\frac{(4q+3q_F)q_F}{2^n - q_F} < 2\frac{(6q+5q_F)q_F}{2^n - 2(q_F+q)q_F}.$$

Now, since

$$|\Pr_{EM}(E^c) - \Pr_{perm}(E^c)| = |(1 - \Pr_{EM}(E)) - (1 - \Pr_{perm}(E))| =$$
$$= |\Pr_{EM}(E) - \Pr_{perm}(E)|,$$

we get that

$$|\Pr_{EM}(E) - \Pr_{perm}(E)| =$$
$$= \frac{1}{2}\left(|\Pr_{EM}(E) - \Pr_{perm}(E)| + |\Pr_{EM}(E^c) - \Pr_{perm}(E^c)|\right) <$$
$$< \frac{(6q+5q_F)q_F}{2^n - 2(q_F+q)q_F}$$

□

3.3 A lower bound for the advantage - Proof of Proposition 2

For the proof of Proposition 2 we need the following technical lemma

Lemma 7. *Let* $EM := EM_{\pi^2_{f_1},\pi^2_{f_2};k_1,k_2,k_3}$. *If* $x, y \in \{0,1\}^{2n}$ *such that*

$$(EM(x))_L = (EM(y))_L$$
$$(x \oplus EM(x))_R = (y \oplus EM(y))_R$$

then $x = y$.

Proof. Denote

$$\check{x} := \pi^2_{f_1}(x \oplus k_1) \oplus k_2,$$
$$\check{y} := \pi^2_{f_1}(y \oplus k_1) \oplus k_2.$$

We have that

$$(EM(x))_R = \check{x}_R \oplus f_2((EM(x))_L) \oplus (k_3)_R = (x \oplus k_1)_R \oplus f_1(\check{x}_L) \oplus f_2((EM(x))_L) \oplus (k_3)_R.$$

Hence

$$f_1(\check{x}_L) = (x \oplus EM(x))_R \oplus f_2((EM(x))_L) \oplus (k_1 \oplus k_3)_R.$$

Similarly

$$f_1(\check{y}_L) = (y \oplus EM(y))_R \oplus f_2((EM(y))_L) \oplus (k_1 \oplus k_3)_R.$$

so since $(EM(x))_L = (EM(y))_L$ and $(x \oplus EM(x))_R = (y \oplus EM(y))_R$ we get that $f_1(\check{x}_L) = f_1(\check{y}_L)$, and then, since f_1 is one-to-one, that $\check{x}_L = \check{y}_L$. Now

$$(EM(x))_L = \check{x}_L \oplus f_2(\check{x}_R) \oplus (k_3)_L,$$

hence

$$f_2(\check{x}_R) = (EM(x))_L \oplus \check{x}_L \oplus (k_3)_L.$$

Similarly

$$f_2(\check{y}_R) = (EM(y))_L \oplus \check{y}_L \oplus (k_3)_L,$$

so since $(EM(x))_L = (EM(y))_L$ and $\check{x}_L = \check{y}_L$ we get that $f_2(\check{x}_R) = f_2(\check{y}_R)$, and then, since f_2 is one-to-one, that $\check{x}_R = \check{y}_R$. Therefore $\check{x} = \check{y}$, i.e.,

$$\pi^2_{f_1}(x \oplus k_1) \oplus k_2 = \pi^2_{f_1}(y \oplus k_1) \oplus k_2$$

hence $x = y$.

□

Proof of Proposition 2. For simplicity, assume q_F is even. Suppose that the adversary uses the following strategy. First he takes arbitrary distinct $a, b \in \{0,1\}^n$. Then he makes $q_F/2$ queries of the form $F(x * a) =?$ for distinct x's in $\{0,1\}^n$ and $q_F/2$ queries of the form $F(y * b) =?$ for distinct y's in $\{0,1\}^n$ (not necessarily different from the x's). The adversary will guess that F is a random permutation if and only if there are x and y for which $(F(x * a))_L = (F(y * b))_L$ and $(F(x * a))_R \oplus a = (F(y * b))_R \oplus b$. The previous lemma guarantees that if F is a BPEM then the adversary will not find such x and y. To conclude the proof, it only remains to show that the probability that the adversary finds such x and y is at least $\frac{1}{8} \cdot \frac{q_F^2}{2^{2n}}$. One way to show this is to observe that by Bonferroni inequality, this probability is at least

$$\frac{(q_F/2)^2}{2^{2n}-1} - \frac{2\binom{q_F/2}{2}^2}{(2^{2n}-1)(2^{2n}-3)}.$$

□

4 A practical constructions: 256-bit cipher EM256AES

We show here a practical construction of a 256-bit block cipher based on the 2-rounds BPEM EM cipher, where the underlying permutation is AES.

Definition 5 ($EM256AES$: a 256-bit block cipher). *Let ℓ_1 and ℓ_2 be two 128-bit keys, selected uniformly and independently at random, from $\{0,1\}^{128}$, and let k_1, k_2, k_3 be three 256-bit secret keys selected uniformly and independently at random from $\{0,1\}^{256}$. Let the permutations f_1 and f_2 be the AES encryption using ℓ_1 and ℓ_2 as the AES key, respectively.*
The 256-bit block cipher $EM256AES$ is defined as the associated instantiation of the 2-rounds BPEM cipher $EM\pi^2_{f_1}, \pi^2_{f_2}, k_1, k_2, k_3$.
Usage:

- *ℓ_1 and ℓ_2 are determined during the setup phase, and can be made public (e.g., sent from the sender to the receiver as an IV).*
- *k_1, k_2, k_3 are selected per encryption session.*

Single key EM256AES is the special case where a single value $k \in \{0,1\}^{256}$ and a single value $\ell \in \{0,1\}^{128}$ are selected uniformly and independently at random, and the EM256AES cipher uses $k_1 = k_2 = k_3 = k$ and $\ell_1 = \ell_2 = \ell$.

Hereafter, we use the single key EM256AES. To establish security properties for $EM256AES$, we make the standard assumption about AES: if a secret key is selected (uniformly at ranomd), an adversary has negligible advantage in distinguishing AES from a random permutation of $\{0,1\}^{128}$ even after seeing a (very) large number of plaintext-ciphertext pairs (i.e., that AES satisfies its design goals ([1], Section 4). This assumption is certainly reasonable if the number of blocks that are encrypted with the same keys setup is limited to be much smaller than 2^{64}. Therefore, in our context, we can consider assigning the randomly selected key ℓ at setup time as an approximation for a random selection of the two (identical here) permutations $f_1 = f_2$. Under this assumption, we can rely on the result of Theorem 4 for the security of $EM256AES$.

***EM256AES* efficiency:** An encryption session between two parties requires exchanging a 256-bit secret key and transmitting a 128-bit IV ($= \ell$). One key (and IV) can be used for N blocks as long as we keep $N \ll 2^{64}$.
Computing one (256-bit) ciphertext involves 4 AES computations (with the IV as the AES key) plus a few much cheaper XOR operations. Let us assume that the encryption is executed on a platform that has the capability of computing AES at some level of performance. If the $EM256AES$ encryption (decryption) is done in a serial mode, we can estimate the encryption rate to be roughly half the rate of AES (serial) computation on that platform (4 AES operations per on 256-bit block). Similarly, if the $EM256AES$ encryption is done in a parallelized mode, we can estimate the throughput to be roughly half the throughput of AES.

***EM256AES* performance:** To test the actual performance of $EM256AES$, and validate our predictions, we coded an optimized implementation of $EM256AES$. Its performance is reported here.
The performance was measured on an Intel Core i7-4700MQ (microarchitecture Codename Haswell) where the enhancements (Intel Turbo Boost Technology, Intel Hyper-Threading Technology, and Enhanced Intel Speedstep Technology) were disabled. The code used the AES instructions (AES-NI) that are available on such modern processors.
On this platform, we point out the following baseline: the performance of AES (128-bit key) in a parallelized mode (CTR) is 0.63 C/B, and in a serial mode (CBC) it is 4.44 cycles per byte (C/B hereafter).
The measured performance of our $EM256AES$ implementation was 1.44 C/B for the parallel mode, and 8.92 C/B for the serial mode. The measured performance clearly matches the predictions.
It is also interesting to compare the performance of $EM256AES$ to another 256-bit cipher. To this end, we prepared an implementation of Rijndael256 cipher [4] [2]. For details on how to code Rijndael256 with AES-NI, see [8]). Rijndael256 (in ECB mode) turned out to be much slower than $EM256AES$, performing at 3.85 C/B.

5 Discussion

We presented here a new variation, BPEM, of the EM cipher, which is immune against attacks that attempt to leverage any non-uniform behavior of $P(x) \oplus x$. Theorem 1 (and its variants, Theorems 2, 3, 4) implies that if $BPEM$ cipher is used much less than $2^{n/2}$ times, the probability of distinguishing it from a random permutation, let alone breaking it in any sense, is negligible.

One interesting feature of our EM-based construction is obtaining a $2n$-bit block size while using an n-bit permutation primitive. The computational cost of encrypting (decrypting) one $2n$-bit block is 4 evaluation of an n-bit permutation (plus a relatively small overhead). Note that this make BEMP ready to be used in practice, for example to define a 256-bit cipher, because "good" permutations of $\{0,1\}^{128}$ are available. We demonstrated the specific cipher $EM256AES$, which is based on AES, and confirmed its efficiency to be (only) half the performance of AES (2.5 times faster than Rijndale256). With the standard assumption on AES, Theorem 4 ensures that it can be used up to 2^{52} times with the same keys before its outputs can be distinguished from a random permutation of $\{0,1\}^{256}$ with an advantage exceeding 10^{-6}. Changing a key/IV every 2^{52} encryptions does not impose a practical limitation.

A variation on the way by which BPEM is used, can make it a naturally tweakable $2n$-bit block cipher as follows. A public IV ($=\ell$) can be associated with each encrypted block as its unique identifier, and can therefore be viewed

[2] AES is based on the Rijndael block cipher. While AES standardizes only a 128 block size, the Rijndael definitions support both 128-bit and 256-bit blocks

as the tweak. To randomized these IV's, it is possible to encrypt (using k) a block identifier such as its "address".

The expression of the advantage Theorem 1 behaves linearly with the number of queries to the public permutations, and quadratically with q_F. This proves the security of a protocol, described below, for using BPEM cipher where the secret keys are changed more frequently than the public permutations. This protocol fits well the reasonable assumption that the essential limitations on the number of adversary queries should be on the encryption/decryption invocations, while no limitations should be imposed on the number of queries to the public permutations.

Protocol: choose the public permutations for a period of $\frac{1}{1000}2^{2n/3}$ blocks, divided into $2^{n/3}$ sessions of $\frac{1}{1000}2^{n/3}$ blocks. Change the secret keys every session.

This way, the relevant information on the block cipher, from a specific choice of keys, is limited to a session, while the adversary can accumulate relevant information from replies to the public permutations across sessions. Therefore, q_F is limited to $\frac{1}{1000}2^{n/3}$, while q is limited to $\frac{1}{1000}2^{2n/3}$. Theorem 1 guarantees that this usage is secure.

We point out that the upper bound of Theorem 1 and the lower bound of Proposition 2 leave the problem of determining the actual order of magnitude of the adversary's advantage widely open. Two observations indicate that Theorem 1 is not tight.

- An examination of the proof of Theorem 1 reveals that the randomness of k_2 does not play any role, and it remains valid even if $k_2 = 0$. It is natural to expect that including a random k_2 improve sthe security of the resulting cipher. The choice where $k_2 = 0$ in BPEM degenerates it to a special form of an EM cipher. In this case, the attack of Daemen [3] applies.
- The result of Bogdanov et al. [2], translated to the $2n$-bit block size, shows that the regular 2-rounds EM cipher is indistinguishable from a random permutation of $\{0,1\}^{2n}$ even with $2^{4n/3}$ queries (and this bound is tight for an adversary with unlimited computational resources). The gap between this number and Theorem 1 can be only partly attributed to the fact that the regular EM construction uses permutations of $\{0,1\}^{2n}$, while BPEM uses only permutations of $\{0,1\}^n$.

References

1. –, Announcing request for candidate algorithm nominations for the Advanced Encryption Standard (AES), http://csrc.nist.gov/CryptoToolkit/aes/preround1/aes_9709.htm (1997).
2. A. Bogdanov, L. R. Knudsen, G. Leander, F-X Standaert, J. P. Steinberger, and E. Tischhauser, Key-alternating ciphers in a provable setting: encryption using a small number of public permutations (extended abstract), in *Advances in cryptology—EUROCRYPT 2012*, Lecture Notes in Comput. Sci., 7237, Springer, Heidelberg, 45–62 (2012).

3. J. Daemen, Limitations of the Even-Mansour construction, in *Advances in cryptology—ASIACRYPT '91*, Lecture Notes in Computer Science, 739, Springer, Berlin, (H. Imai, R. L. Rivest, T. Matsumoto, editors) 495-498 (1993).
4. J. Daemen, V. Rijmen, AES Proposal: Rijndael (National Institute of Standards and Technology), http://csrc.nist.gov/archive/aes/rijndael/Rijndael-ammended.pdf (1999).
5. O. Dunkelman , N. Keller, A. Shamir, Minimalism in Cryptography: The Even-Mansour Scheme Revisited, in *Advances in cryptology—EUROCRYPT 2012*, Lecture Notes in Comput. Sci., 7237, Springer, Heidelberg. 336–354 (2012).
6. I. Dinur, O. Dunkelman, N. Keller, A. Shamir, Key Recovery Attacks on 3-round Even-Mansour, 8-step LED-128, and full AES^2, http://eprint.iacr.org/2013/391 (2013).
7. S. Even, Y. Mansour, A construction of a cipher from a single pseudorandom permutation, J. Cryptology **10** (1997), no. 3, 151–161.
8. S. Gueron, Intel Advanced Encryption Standard (AES) Instructions Set (Rev 3.01), http://software.intel.com/sites/default/files/article/165683/aes-wp-2012-09-22-v01.pdf (2014).
9. C. Hall, D. Wagner, J. Kelsey, B. Schneier, Building PRFs from PRPs, in: Proceedings of CRYPTO-98: Advances in Cryptography, Springer Verlag, 1998, pp. 370-389.
10. I. Nikolić, L. Wang, S. Wu, Cryptanalysis of Round-Reduced LED. In *FSE*, 2013. To appear in Lecture Notes in Computer Science.