# A Security Analysis of Key Expansion Functions Using Pseudorandom Permutations

Ju-Sung Kang, Nayoung Kim, Wangho Ju, Ok-Yeon Yi



# A Security Analysis of Key Expansion Functions Using Pseudorandom Permutations

Ju-Sung Kang[1], Nayoung Kim[2], Wangho Ju[2], Ok-Yeon Yi[1]

[1] Dept. of Mathematics, Kookmin University
[2] Dept. of Financial Information Security, Graduate School, Kookmin University
77 Jeongneung-Ro, Seongbuk-Gu, Seoul, 136-702, KOREA
{jskang,izerotwo,nandars2,oyyi}@kookmin.ac.kr

**Abstract.** Within many cryptographic systems a key expansion function is used in order to derive more keying material from the master secret. The derived additional keys may be needed for multiple entities or for different cryptographic purposes such as privacy and authenticity. In this paper we wish to examine the soundness of the key expansion functions on the view point of provable security framework. Especially we focus on the key expansion functions using PRFs(pseudorandom functions) which are recommended by NIST, and show that the variant of Double-Pipeline Iteration mode using PRPs(pseudorandom permutations) is secure, while the variants of Counter and Feedback modes using PRPs are insecure. In practice secure block ciphers such as AES can be regarded as PRPs.

**Keywords:** Privacy, Authenticity, Key expansion function, Keying material, Provable security, Pseudorandomness, PRF, PRP.

## 1 Introduction

Cryptographic keys are essential to the security of all cryptographic algorithms and protocols for some information security objectives, such as privacy or confidentiality, authenticity, digital signature, and non-repudiation in the presence of adversaries. Key management is an indispensable part of the cryptographic system and this includes dealing with the generation, exchange, storage, use, and replacement of keys. Cryptographic systems may use some kinds of keys more than one master key, so key derivation mechanism is contained in the key management part. Key derivation mechanism derives one or more secret keys from a shared secret such as a master key. Thus any key derivation mechanism has the key expansion step. For example, NIST SP 800-56C[10] specifies a key derivation mechanism that is an extraction-then-expansion procedure. This procedure consists of a randomness extraction step and a key expansion step. The randomness extraction step outputs a key derivation key from a master key. A key derivation key is then used as input to the key expansion step that derives keying material and can also be used to derive more keying material from derived keys of key expansion step. The derived additional keys from a key expansion step may be

used for multiple entities or for different cryptographic objectives. Our research interest is to examine the soundness of some key expansion functions (KEFs) for key expansion steps on the view point of provable security framework.

### 1.1 Related Work

Although a key derivation mechanism has the central role in applied cryptography, there has been relatively little formal work addressing the design and security analysis. Krawczyk[6] associated the notion of cryptographically strong secret keys with that of pseudorandom keys, namely, indistinguishable by feasible computation from a uniformly distributed string of the same length, and provided detailed rationale for the hash-based design of key derivation mechanisms based on the extract-then-expand approach. The extraction step generates a uniformly random or pseudorandom seed key from the master key that may be an output of an imperfect physical random number generator, and the expansion step derives several additional pseudorandom cryptographic keys from the seed key.

Gilbert[4] investigated the security of block cipher modes of operation allowing to expand an one-block input into a longer $t$-block output, under the Luby-Rackoff security paradigm[7, 8] which is originally due to the indistinguishability in Goldreich-Goldwasser-Micali[5]. A KEF in a key derivation mechanism is a typical example of the one-block-to-many modes of operation. In [4], the author showed that, under the Luby-Rackoff security model, the key expansion function MILENAGE of 3GPP[13] is pseudorandom.

On the other hand NIST has specified three KEFs using pseudorandom functions (PRFs) in SP 800-108[11]. A PRF family $\{PRF_s(\cdot)|s \in S\}$ consists of polynomial time computable functions with an index $s$, a seed, such that when $s$ is randomly selected from $S$ and not known to adversaries, $PRF_s(\cdot)$ is computationally indistinguishable from a random function defined on the same domain and range [5]. In [11], several families of PRF-based key expansion functions are defined without describing the internal structure of the PRF, and recommended the use of either HMAC[3] or CMAC[9] as the PRF.

### 1.2 Our Contribution

In spite of several years after the publication of NIST SP 800-108[11], as far as we know there is no noticeable result that deals with a security analysis for three KEFs of this document. It seems that if a PRF, such as HMAC or CMAC, is used as the building block of the three KEFs in [11], we have difficulty in investigating the soundness of the given schemes. Hence we add a constraint condition that a pseudorandom permutation (PRP), such as AES, is used as the building block of the given KEFs. A PRP family is a special case of PRF families and computationally indistinguishable from a random permutation defined on the same domain. Once we regard the given KEFs of [11] as PRP-based schemes, we can investigate the security of these variant schemes in the Luby-Rackoff security model which is similar to the context of Gilbert[4].

In fact NIST SP 800-108[11] defines three families of PRF-based KEFs, so called, Counter mode, Feedback mode and Double-Pipeline Iteration mode. In this work we consider the variant schemes of three KEFs that use PRPs, and show that the variant of Double-Pipeline Iteration mode using PRPs is pseudorandom, while the variants of Counter and Feedback modes using PRPs are insecure. Moreover we provide a concrete security bound for the variant of Double-Pipeline Iteration mode where the underlying PRP is a practical block cipher. This concrete security approach is based on the security model of Bellare-Killian-Rogaway[1] and Bellare-Rogaway[2].

## 2 Notions of PRF and PRP

In order to examine the soundness of some KEFs, we need to introduce the rigorous notions of PRF and PRP. These are useful conceptual starting points to enable the security analysis in the design of some cryptographic functions. Cryptographic functions such as block ciphers or their modes of operation can be regarded as a pseudorandom function family indexed by a uniformly distributed key space. It is natural that we also consider a KEF as a pseudorandom function family because it is an example of block cipher modes of operation where a PRP is used as an underlying primitive in the KEF. We have to recognize that a KEF is an instance of a PRF to obtain the theoretical upper bound for the provable security. No computationally efficient adversary can distinguish with significant advantage between a randomly chosen instance of a PRF and a uniformly selected random function of the same domain and range. In this section we describe concrete security approach which is based on the Bellare-Rogaway[2] security model.

### 2.1 Function Families

A function family is a map $\Lambda : \mathcal{K} \times D \to R$, where $\mathcal{K}$ is the keyspace, $D$ is the domain and $R$ is the range of $\Lambda$. The two-input function $\Lambda$ takes a key $K$ and an input $x$ to return a point $y$ we denote by $\Lambda(K, x)$. For any key $K \in \mathcal{K}$ we define the map $\Lambda_K : D \to R$ by $\Lambda_K(x) = \Lambda(K, x)$. We call the function $\Lambda_K$ an instance of the function family $\Lambda$. Thus $\Lambda$ specifies a collection of maps indexed by the key space. Usually the probability distribution of a function family comes from some probability distributions on the keyspace $\mathcal{K}$. Unless otherwise indicated, this distribution will be the uniform distribution.

We use the following notation in this paper. For any positive integer $k$, $n$ and $m$, we denote $\mathcal{K} = \{0,1\}^k, D = \{0,1\}^n$ and $R = \{0,1\}^m$, where $k$, $n$, and $m$ are called the key-length, the input-length and the output-length, respectively. We denote by $K \xleftarrow{\$} \mathcal{K}$ the operation of selecting a random string $K$ from $\mathcal{K}$. The notation $f \xleftarrow{\$} \Lambda$ means the operation $K \xleftarrow{\$} \mathcal{K}$ and $f = \Lambda_K$. In other words, let $f$ be the function $\Lambda_K$ where $K$ is a randomly chosen key. We are interested in the input-output behavior of this randomly chosen instance of the family.

There are two particular function families that we need to consider in order to define PRFs and PRPs. One is $\mathcal{F}(D, R)$ the family of all functions from domain $D$ to range $R$, the other is $\mathcal{P}(D)$ the family of all permutations on $D$. A uniformly chosen instance of $\mathcal{F}(D, R)$ is called a *random function* from $D$ to $R$, and a uniformly chosen instance of $\mathcal{P}(D)$ is called a *random permutation* on $D$. The key describing any particular instance function is simply a description of this instance function in some canonical notation. For example, order the domain $D$ lexicographically as $x_1, x_2, \ldots$, and let the key for a function $f$ be the list of values $(f(x_1), f(x_2), \ldots)$. The keyspace of $\mathcal{F}(D, R)$ is simply the set of all these keys, under the uniform distribution. The key for a function in this family is a list of all the output values of the function as its input ranges over $\{0,1\}^n$. Namely, the key describing a particular instance function is exactly corresponding to the function itself. Note that the size of the key spaces of $\mathcal{F}(D, R)$ and $\mathcal{P}(D)$ are $(2^m)^{2^n}$ and $(2^n)!$, respectively.

### 2.2 Pseudorandom Functions and Permutations

A *pseudorandom function* is a function family with the property that the input-output behavior of a random instance of the family is computationally indistinguishable from that of a random function. Similarly, a function family is a *pseudorandom permutation* if the input-output behavior of a random instance of the family is computationally indistinguishable from that of a random permutation. In order to introduce the notions of PRF and PRP, we consider the following security model. The notion of PRP is very similar to the one of PRF. Thus we only consider the notion of PRF. Let any adversary $\mathcal{A}$ be an algorithm to distinguish a random instance of a function family from a random function. The adversary $\mathcal{A}$ has access to an oracle. The oracle will be chosen either as a random instance of a function family or as a random function by coin tossing. When the oracle selects a fuction as $G$ we consider two different worlds. Usually in World 0, $G$ will be chosen as a random function, while in World 1, $G$ will be chosen as a random instance of a function family. And the adversary must determine in which world it is placed, and at the end of its computation outputs a bit.

In the formalization, we consider two different ways in which $G$ will be chosen, giving rise to two different worlds.

**World 0.** The function $G$ is drawn at random from $\mathcal{F}(D, R)$, namely, the function $G$ is selected via $G \xleftarrow{\$} \mathcal{F}(D, R)$ .

**World 1.** The function $G$ is drawn at random from $\Lambda$, namely, the function $G$ is selected via $G \xleftarrow{\$} \Lambda$.

**Definition 1.** *Let $\Lambda : \mathcal{K} \times D \to R$ be a function family, and $\mathcal{A}$ be an algorithm that takes an oracle for a function $G : D \to R$, and returns a bit. We consider two experiments:*

| *Experiment* $\mathbf{Exp}_F^{prf-1}(\mathcal{A})$ | *Experiment* $\mathbf{Exp}_F^{prf-0}(\mathcal{A})$ |
|---|---|
| $K \xleftarrow{\$} \mathcal{K}$ | $G \xleftarrow{\$} \mathcal{F}(D,R)$ |
| $b \xleftarrow{\$} \mathcal{A}^{\Lambda_K}$ | $b \xleftarrow{\$} \mathcal{A}^{G}$ |
| *Return* $b$ | *Return* $b$ |

*The prf-advantage of* $\mathcal{A}$ *is defined by*

$$\begin{aligned}\mathbf{Adv}_{\Lambda}^{prf}(\mathcal{A}) &= Pr\left[\mathbf{Exp}_F^{prf-1}(\mathcal{A}) = 1\right] - Pr\left[\mathbf{Exp}_F^{prf-0}(\mathcal{A}) = 1\right] \\ &= Pr\left[\mathcal{A}^G = 1 \mid G \leftarrow \Lambda\right] - Pr\left[\mathcal{A}^G = 1 \mid G \leftarrow \mathcal{F}(D,R)\right] \ .\end{aligned}$$

## 3 PRP-based KEFs of NIST SP 800-108

NIST has specified three KEFs using PRFs in SP 800-108[11] without describing the internal structure of the PRF, and recommended the use of either HMAC[3] or CMAC[9] as the underlying PRF. However we have difficulty in analyzing the provable security of the given KEFs where the underlying primitives are PRFs. Thus we change this PRF condition to the PRP one because a PRP family is a special case of PRF families. That is, hereafter we consider only PRP-based KEFs of the NIST recomendations. Once we regard the given KEFs of [11] as PRP-based schemes, we can investigate the provable security of these variant schemes in the Luby-Rackoff security model which is similar to the context of Gilbert[4].

### 3.1 PRP-based Counter mode

Let $\mathcal{G}$ be a PRP on $\{0,1\}^n$, then the PRP-based Counter mode is defined in Definition 2 and illustrated in Figure 1.

**Definition 2.** *For any permutation* $g \in \mathcal{G}$ *and integer* $t \geq 2$, $CNT[g]$ *is called a PRP-based Counter mode if*

$$\begin{aligned}&CNT[g] : \{0,1\}^n \to \{0,1\}^{nt} \ , \\ &CNT[g](x) = (z_1, z_2, \ldots, z_t) = (g(x \oplus c_1), g(x \oplus c_2), \ldots, g(x \oplus c_t)) \ ,\end{aligned}$$

*where* $c_1, c_2, \ldots, c_t$ *are constants of* $\{0,1\}^n$.

### 3.2 PRP-based Feedback mode

PRP-based Feedback mode is defined in Definition 3 and illustrated in Figure 2.

**Definition 3.** *For any permutation* $g \in \mathcal{G}$ *and integer* $t \geq 2$, $FB[g]$ *is called a PRP-based Feedback mode if*

$$\begin{aligned}&FB[g] : \{0,1\}^n \to \{0,1\}^{nt} \ , \\ &FB[g](x) = (z_1, z_2, \ldots, z_t) = (g(x \oplus c_1), g(x \oplus z_1 \oplus c_2), \ldots, g(x \oplus z_{t-1} \oplus c_t)) \ ,\end{aligned}$$

*where* $c_1, c_2, \ldots, c_t$ *are constants of* $\{0,1\}^n$.

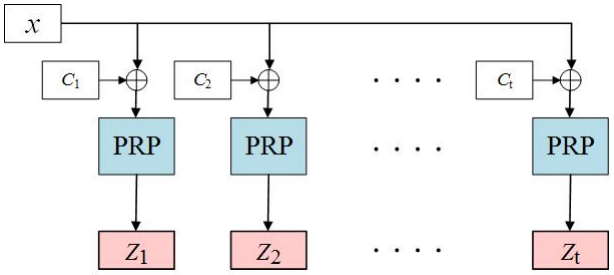


**Fig. 1.** PRP-based Counter mode

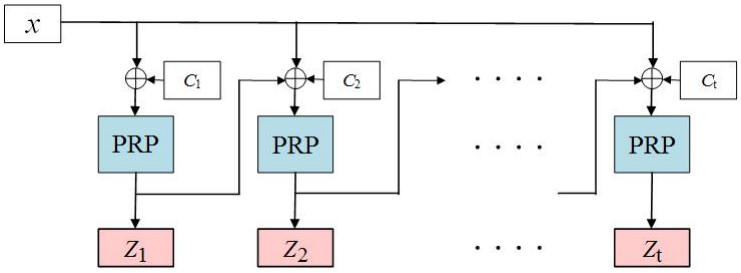


**Fig. 2.** PRP-based Feedback mode

### 3.3 PRP-based Double-Pipeline Iteration mode

PRP-based Double-Pipeline Iteration mode is defined in Definition 4 and illustrated in Figure 3.

**Definition 4.** *For any permutation $g \in \mathcal{G}$ and integer $t \geq 2$, $DP[g]$ is called PRP-based Double-Pipeline Iteration mode if*

$$\begin{aligned} DP[g] &: \{0,1\}^n \to \{0,1\}^{nt}, \\ DP[g](x) &= (z_1, z_2, \ldots, z_t) \\ &= \left(g\left(g(x) \oplus x \oplus c_1\right), g\left(g^2(x) \oplus x \oplus c_2\right), \ldots, g\left(g^t(x) \oplus x \oplus c_t\right)\right), \end{aligned}$$

*where for each $1 \leq k \leq t$, $g^k$ denotes $k$ times iteration of $g$ and $c_1, c_2, \ldots, c_t$ are constants of $\{0,1\}^n$.*

## 4 Provable Security of PRP-based KEFs

In this section we show that the PRP-based Double-Pipeline Iteration mode is secure, while the PRP-based Counter mode and Feedback mode are insecure. Since a secure block cipher, such as AES, is regarded as the underlying PRP in practice, the PRP-based Double-Pipeline Iteration mode can be recommended

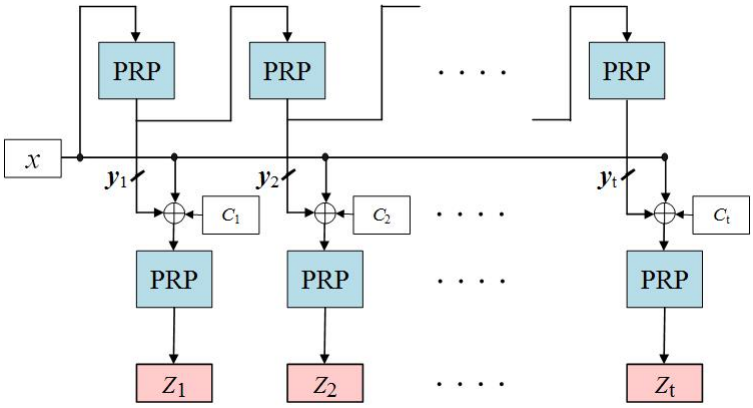


**Fig. 3.** PRP-based Double-Pipeline Iteration mode

as a candidate of secure practical KEFs using block ciphers. On the other hand Gilbert[4] already showed that the functions associated with the counter mode and the OFB mode are not pseudorandom. The counter mode of [4] is the same as PRP-based Counter mode in Definition 2, but the OFB mode of [4] is slightly different from PRP-based Feedback mode in Definition 3. In this work we propose somewhat different and more clear processes of proving the insecurities of PRP-based Counter and Feedback modes.

### 4.1 Insecurity of PRP-based Counter mode

Now we consider the case that $CNT[\pi]$ is derived from a random permutation $\pi \in \mathcal{P}(\{0,1\}^n)$. Then it is simple that $CNT[\pi]$ is insecure.

**Theorem 1.** *For any random permutation $\pi \in \mathcal{P}$, $CNT[\pi]$ is not a secure PRF.*

***Proof.*** In order to show that $CNT[\pi]$ is not a secure PRF we specify an adversary attacking $CNT[\pi]$. Since an instance of $CNT[\pi]$ is a function from $\{0,1\}^n$ to $\{0,1\}^{nt}$, the adversary $\mathcal{A}$ will get an oracle for a function $G$ that maps $\{0,1\}^n$ to $\{0,1\}^{nt}$. In World 0, $G$ will be chosen as a random function from $\mathcal{F} = \mathcal{F}(\{0,1\}^n, \{0,1\}^{nt})$, while in World 1, $G$ will be set to $CNT[\pi]$ where $\pi$ is a random permutation from $\mathcal{P}(\{0,1\}^n)$. The adversary $\mathcal{A}$ must determine in which world it is placed. Let us show how the adversary $\mathcal{A}$ works.

Adversary $\mathcal{A}^G$
$(z_1^{(1)}, \ldots, z_t^{(1)}) \leftarrow G\left(x^{(1)}\right)$
$(z_1^{(2)}, \ldots, z_t^{(2)}) \leftarrow G\left(x^{(2)} = x^{(1)} \oplus c_1 \oplus c_2\right)$
**if** $z_1^{(1)} = z_2^{(2)}$ and $z_2^{(1)} = z_1^{(2)}$ **then return** 1
**else return** 0

If $G = CNT[\pi]$ for some $\pi$, it is certainly true that $z_1^{(1)} = z_2^{(2)}$ and $z_2^{(1)} = z_1^{(2)}$. On the other hand if $G$ is a random function from $\mathcal{F}$, the probability of the event that $z_1^{(1)} = z_2^{(2)}$ and $z_2^{(1)} = z_1^{(2)}$ will be $2^{-2n}$, the probability that $\mathcal{A}$ will return 1. Therefore the advantage of $\mathcal{A}$ is as follows:

$$\begin{aligned}\mathbf{Adv}^{prf}_{CNT[\pi]}(\mathcal{A}) &= Pr\left[\mathbf{Exp}^{prf-1}_{CNT[\pi]}(\mathcal{A}) = 1\right] - Pr\left[\mathbf{Exp}^{prf-0}_{CNT[\pi]}(\mathcal{A}) = 1\right] \\ &= Pr\left[\mathcal{A}^G = 1 \mid G \leftarrow CNT[\pi]\right] - Pr\left[\mathcal{A}^G = 1 \mid G \leftarrow \mathcal{F}\right] \\ &= 1 - 2^{-2n}.\end{aligned}$$

From the above formula we obtain that there exists an extremely efficient adversary whose prf-advantage against $CNT[\pi]$ is almost one. This means that $CNT[\pi]$ is not a secure PRF. $\square$

### 4.2 Insecurity of PRP-based Feedback mode

We investigate the provable security of the PRP-based Feedback mode $FB[\pi]$ derived from a random permutation $\pi$. It is also a simple argument similar to the case of PRP-based Counter mode that $FB[\pi]$ is insecure.

**Theorem 2.** *For any random permutation $\pi \in \mathcal{P}$, $FB[\pi]$ is not a secure PRF.*

***Proof.*** We find an adversary with a high advantage attacking $FB[\pi]$ in order to prove that $FB[\pi]$ is not a secure PRF. Since an instance of $FB[\pi]$ is a function from $\{0,1\}^n$ to $\{0,1\}^{nt}$, the adversary $\mathcal{A}$ will get an oracle for a function $G$ that maps $\{0,1\}^n$ to $\{0,1\}^{nt}$. In World 0, $G$ will be chosen as a random function of $\mathcal{F} = \mathcal{F}(\{0,1\}^n, \{0,1\}^{nt})$, while in World 1, $G$ will be set to $FB[\pi]$ where $\pi$ is a random permutation from $\mathcal{P}(\{0,1\}^n)$. The adversary $\mathcal{A}$ must determine in which world it is placed. In this case $\mathcal{A}$ queries its oracle at the $x^{(1)}$ to get back $(z_1^{(1)}, \ldots, z_t^{(1)})$ and then queries its oracle at the $x^{(2)}$ to get back $(z_1^{(2)}, \ldots, z_t^{(2)})$. The adversary $\mathcal{A}$ works as follows:

Adversary $\mathcal{A}^G$
$(z_1^{(1)}, \ldots, z_t^{(1)}) \leftarrow G\left(x^{(1)}\right)$
$(z_1^{(2)}, \ldots, z_t^{(2)}) \leftarrow G\left(x^{(2)} = x^{(1)} \oplus c_1 \oplus c_2 \oplus z_1^{(1)}\right)$
**if** $z_1^{(2)} = z_2^{(1)}$ **then return** 1
**else return** 0

If $G = FB[\pi]$ for some $\pi$, it is obvious that

$$z_1^{(2)} = \pi((x^{(1)} \oplus c_1 \oplus c_2 \oplus z_1^{(1)}) \oplus c_1) = \pi(x^{(1)} \oplus c_2 \oplus z_1^{(1)}) = z_2^{(1)}.$$

On the other hand if $G$ is a random function, the probability that $z_1^{(2)} = z_2^{(1)}$ will be $2^{-n}$, the probability that $\mathcal{A}$ will return 1. Thus $\mathbf{Adv}^{prf}_{FB[\pi]}(\mathcal{A}) = 1 - 2^{-n}$, this shows that $FB[\pi]$ is not a secure PRF. $\square$

### 4.3 Provable security of PRP-based Double-Pipeline Iteration mode

Now we examine the provable security of the PRP-based Double-Pipeline Iteration mode. We consider the case where $DP[\pi]$ is derived from a random permutation $\pi$ and prove that $DP[\pi]$ is a secure PRF. At first we introduce a very useful fact of [12] and [4] for obtaining an upper bound on $\mathbf{Adv}^{prf}(\mathcal{A})$ based on the transition probability $Pr(\mathbf{x}\rightarrow\mathbf{y})$.

**Proposition 1.** *Let $E$ be a randomly chosen instance of a function family $\Lambda$ with the domain $\{0,1\}^n$ and the range $\{0,1\}^m$, $F$ be a random function from $\mathcal{F}=\mathcal{F}(\{0,1\}^n,\{0,1\}^m)$ and $q$ be an integer. An adversary $\mathcal{A}$ will get an oracle for a function $G$. In World 0, $G$ will be chosen from $\mathcal{F}$, while in World 1, $G$ will be set to $E$ that is a randomly drawn from $\Lambda$. The adversary must determine in which world it is placed. Denote by $X$ the subset of $(\{0,1\}^n)^q$ containing all pairwise distinct $q$-tuples $\mathbf{x}=(x^{(1)},\cdots,x^{(q)})$. If there exist a subset $Y$ of $(\{0,1\}^m)^q$ and two positive real numbers $\varepsilon_1$ and $\varepsilon_2$ such that*

$$(a)\ |Y| \geq (1-\varepsilon_1)\cdot 2^{mq}$$
$$(b)\ \text{for each } \mathbf{x}\in X \text{ and } \mathbf{y}\in Y,\quad Pr\left(\mathbf{x} \overset{E}{\rightarrow} \mathbf{y}\right) \geq (1-\varepsilon_2)\cdot\frac{1}{2^{mq}}\ ,$$

*then for any adversary $\mathcal{A}$ using $q$ queries*

$$\begin{aligned}\mathbf{Adv}^{prf}_{\Lambda}(\mathcal{A}) &= Pr\left[\mathcal{A}^G=1|G\leftarrow\Lambda\right]-Pr\left[\mathcal{A}^G=1|G\leftarrow\mathcal{F}\right]\\ &\leq \varepsilon_1+\varepsilon_2\ .\end{aligned}$$

By the argument using Proposition 1, in the following Theorem 3 we obtain an upper bound on $\mathbf{Adv}^{prf}_{DP[\pi]}(\mathcal{A})$. From this we know the fact that the PRP-based Double-Pipeline Iteration mode is a secure PRF.

**Theorem 3.** *For any PRP-based Double-Pipeline Iteration mode*

$$DP[\pi]:\{0,1\}^n\rightarrow\{0,1\}^{nt}\ ,\quad \forall\pi\in\mathcal{P}(\{0,1\}^n)\ ,$$

*let $\mathcal{A}$ be an adversary with $q$ queries such that $\frac{t^2q^2}{2^n}\leq\frac{2}{3}$. Then we obtain that*

$$\mathbf{Adv}^{prf}_{DP[\pi]}(\mathcal{A})\leq\frac{7t^2q^2}{2^{n+1}}\ .$$

***Proof.*** Let $X$ denote the subset of $(\{0,1\}^n)^q$ containing all pairwise distinct $q$-tuples $\mathbf{x}=(x^{(1)},\ldots,x^{(q)})$ and $Z$ be the set of $q$-tuples $\mathbf{z}=((z_1^{(1)},\ldots,z_t^{(1)}),(z_1^{(2)},\ldots,z_t^{(2)}),\ldots,(z_1^{(q)},\ldots,z_t^{(q)}))\in(\{0,1\}^{nt})^q$, where all $tq$ values of $z_k^{(i)}$, $1\leq k\leq t$ and $1\leq i\leq q$, are distinct. By Proposition 1, it suffices to show that there exist positive real numbers $\varepsilon_1$ and $\varepsilon_2$ such that

$$|Z|\geq(1-\varepsilon_1)\cdot 2^{ntq} \tag{1}$$

and

$$\forall \mathbf{x} \in X,\ \forall \mathbf{z} \in Z, \quad Pr\left(\mathbf{x} \overset{DP[\pi]}{\longrightarrow} \mathbf{z}\right) \geq (1-\varepsilon_2) \cdot \frac{1}{2^{ntq}} \ . \tag{2}$$

Note that

$$\begin{aligned}
\frac{1}{2^{ntq}} \cdot |Z| &= \frac{2^n \cdot (2^n-1) \cdots (2^n-(tq-1))}{2^{ntq}} \\
&= 1 \cdot \left(1-\frac{1}{2^n}\right) \cdots \left(1-\frac{tq-1}{2^n}\right) \\
&\geq 1 - \frac{1}{2^n}\left(1+2+\cdots+(tq-1)\right) \\
&= 1 - \frac{1}{2^n}\left(\frac{(tq-1)\,(1+tq-1)}{2}\right) \\
&\geq 1 - \frac{t^2q^2}{2^{n+1}} \ .
\end{aligned}$$

Then the inequality (1) is established, if we set $\varepsilon_1 = \frac{t^2q^2}{2^{n+1}} > 0$.

In order to estimate the transition probability of (2), we have to consider somewhat complicated cases changed by some input-output conditions associated with $\pi$'s of $DP[\pi]$. For any fixed $\mathbf{x} \in X \subset (\{0,1\}^n)^q$ and $\mathbf{z} \in Z \subset (\{0,1\}^{nt})^q$, the transition probability $Pr\left(\mathbf{x} \overset{DP[\pi]}{\longrightarrow} \mathbf{z}\right)$ can be estimated by investigating the intermediate value $\mathbf{y} \in (\{0,1\}^{nt})^q$ of $DP[\pi]$ depicted in Figure 3. The values of the input $\mathbf{x}$ and the corresponding output $\mathbf{z}$ are known, while the exact value of the intermediate value $\mathbf{y}$ is unknown. Hence we have to collect all possible candidates about the unknown intermediate value. Let $Y \subset (\{0,1\}^{nt})^q$ be the set of all these possible candidates;

$$\begin{aligned}
Y = \Big\{ \mathbf{y} = ((y_1^{(1)}, \ldots, y_t^{(1)}), \ldots, (y_1^{(q)}, \ldots, y_t^{(q)})) \\
\mid\ y_k^{(i)} \neq y_k^{(j)},\ 1 \leq k \leq t,\ 1 \leq i \neq j \leq q \Big\},
\end{aligned}$$

where for each $1 \leq k \leq t$ and $1 \leq i \leq q$, $y_k^{(i)} = \pi^k(x^{(i)})$. By the definition of $Y$,

$$|Y| = 2^{nt} \cdot (2^n-1)^t \cdots (2^n-q+1)^t = \left(\frac{2^n!}{(2^n-q)!}\right)^t \ . \tag{3}$$

Now we introduce a subset $Y' \subset Y$ for convenience of counting distinct input values of $\pi$. Let $Y' = A \cap B \cap C$ with

$$\begin{aligned}
A &= \{\mathbf{y} \in Y \mid y_k^{(i)} \neq y_l^{(j)},\ 1 \leq i,j \leq q,\ 0 \leq k,l \leq t-1\}, \\
B &= \{\mathbf{y} \in Y \mid y_k^{(i)} \neq y_l^{(j)} \oplus x^{(j)} \oplus c_l,\ 1 \leq i,j \leq q\ , 0 \leq k \leq t-1, 1 \leq l \leq t\}, \\
&\quad \text{where } y_0^{(i)} = x^{(i)}, \text{ and} \\
C &= \{\mathbf{y} \in Y \mid y_k^{(i)} \oplus x^{(i)} \oplus c_k \neq y_l^{(j)} \oplus x^{(j)} \oplus c_j,\ 1 \leq i,j \leq q,\ 1 \leq k,l \leq t\}.
\end{aligned}$$

For any $1 \le i \le q$, $1 \le k \le t-1$ and $1 \le l \le t$ , we have

$$\begin{aligned} & Pr\left(\mathbf{x} \overset{DP[\pi]}{\longrightarrow} \mathbf{z}\right) \\ &= \sum_{\mathbf{y}\in Y} Pr\left(x^{(i)} \overset{\pi}{\to} y_1^{(i)},\ y_k^{(i)} \overset{\pi}{\to} y_{k+1}^{(i)}\ ,\ y_l^{(i)} \oplus x^{(i)} \oplus c_l \overset{\pi}{\to} z_l^{(i)}\right) \\ &\ge \sum_{\mathbf{y}\in Y'} Pr\left(x^{(i)} \overset{\pi}{\to} y_1^{(i)},\ y_k^{(i)} \overset{\pi}{\to} y_{k+1}^{(i)}\ ,\ y_l^{(i)} \oplus x^{(i)} \oplus c_l \overset{\pi}{\to} z_l^{(i)}\right) \\ &= |Y'| \cdot \frac{(2^n - 2tq)!}{2^n!}\ . \end{aligned} \tag{4}$$

In order to obtain a lower bound on $|Y'|$, we count $|Y - A|$, $|Y - B|$ and $|Y - C|$. At the first step we represent the set $Y - A$ with three subsets to count $|Y - A|$. For any element of $Y - A$, we consider three cases according to the condition of indexes.

<u>*Case 1.*</u> If $k = 0$ and $1 \le l \le t-1$, then for any $\mathbf{y} \in Y - A$ satisfies $x^{(i)} = y_0^{(i)} = y_l^{(j)}$. In this case the number of $\mathbf{y}$ such that $x^{(i)} = y_l^{(j)}$ is $\frac{|Y|}{2^n} \cdot q^2 \cdot (t-1)$, since by (3), for any fixed $i$, $j$ and $l$, $(2^n)^{t-1} \cdot (2^n - 1)^t \cdots (2^n - q + 1)^t = \frac{|Y|}{2^n}$.

<u>*Case 2.*</u> If $1 \le k \ne l \le t-1$ and $1 \le i = j \le q$, then for any $\mathbf{y} \in Y - A$ satisfies $y_k^{(i)} = y_l^{(j)} = y_l^{(i)}$. In this case the number of $\mathbf{y}$ such that $y_k^{(i)} = y_l^{(i)}$ is $\frac{|Y|}{2^n} \cdot q \cdot \frac{(t-1)(t-2)}{2}$, since by (3), for any fixed $i$, $k$ and $l$, $(2^n)^{t-2} \cdot (2^n - 1)^t \cdots (2^n - q + 1)^t = \frac{|Y|}{(2^n)^2}$.

<u>*Case 3.*</u> If $1 \le k \ne l \le t-1$ and $1 \le i \ne j \le q$, then for any $\mathbf{y} \in Y - A$ satisfies $y_k^{(i)} = y_l^{(j)}$. In this case the number of $\mathbf{y}$ such that $y_k^{(i)} = y_l^{(j)}$ is $\frac{|Y|}{2^n} \cdot \frac{q(q-1)}{2} \cdot \frac{(t-1)(t-2)}{2}$, since by (3), for any fixed $i$, $j$, $k$ and $l$, $(2^n)^{t-2} \cdot (2^n - 1)^t \cdots (2^n - q + 1)^t = \frac{|Y|}{(2^n)^2}$.

Therefore we obtain that

$$|Y - A| \le \frac{|Y|}{2^n} \cdot q \cdot (t-1)\left(q + \frac{t-2}{2} + \frac{(q-1)(t-2)}{4}\right)\ .$$

By the similar argument, we have

$$|Y - B| \le \frac{|Y|}{2^n} \cdot q\left(t + (q-1)t + (t-1)^2 + (q-1)(t-1)^2 + 2(q-1)(t-1)\right)$$

and

$$|Y - C| \le \frac{|Y|}{2^n} \cdot \frac{qt}{2}\left((t-1) + \frac{(q-1)(t-1)}{2} + 2(q-1)\right)\ .$$

Now we have a lower bound on $|Y'|$ as follows:

$$\begin{aligned}|Y'| &\geq |Y| - (|Y - A| + |Y - B| + |Y - C|) \\ &\geq |Y| - \frac{|Y|}{2^{n+1}} q \left(3qt^2 + 4qt - 3q + t^2 - 8t + 5\right) \\ &= |Y| \left(1 - \frac{q}{2^{n+1}} \left(3qt^2 + 4qt - 3q + t^2 - 8t + 5\right)\right). \end{aligned} \tag{5}$$

By (3), (4) and (5), we obtain that

$$\begin{aligned} & Pr\left(\mathbf{x} \overset{DP[\pi]}{\longrightarrow} \mathbf{z}\right) \\ \geq\ & \left(1 - \frac{q}{2^{n+1}} \left(3qt^2 + 4qt - 3q + t^2 - 8t + 5\right)\right) \cdot \frac{(2^n - 2tq)!}{2^n!} \cdot \left(\frac{2^n!}{(2^n - q)!}\right)^t \\ \geq\ & \left(1 - \frac{q}{2^{n+1}} \left(3qt^2 + 4qt - 3q + t^2 - 8t + 5\right)\right) \cdot \frac{1}{2^{ntq}} \left(1 + \frac{qt(3qt-1)}{2^{n+1}}\right) \\ =\ & \frac{1}{2^{ntq}}(1 - \eta)(1 + \delta), \end{aligned}$$

where $\eta = \frac{q}{2^{n+1}} \left(3qt^2 + 4qt - 3q + t^2 - 8t + 5\right)$, $\delta = \frac{qt(3qt-1)}{2^{n+1}}$. Since $\delta \leq 1$ by assumtion $\frac{t^2q^2}{2^n} \leq \frac{2}{3}$,

$$\begin{aligned}(1 - \eta)(1 + \delta) &\geq 1 - 2\eta + \delta \\ &= 1 - \frac{q}{2^{n+1}} \left(3qt^2 + 8qt - 6q + 2t^2 - 15t + 10\right) \\ &\geq 1 - \frac{q}{2^{n+1}} \left(3qt^2 + 3qt^2\right) = 1 - \frac{6q^2t^2}{2^{n+1}}. \end{aligned}$$

Thus

$$Pr\left(\mathbf{x} \overset{DP[\pi]}{\longrightarrow} \mathbf{z}\right) \geq \frac{1}{2^{ntq}} \left(1 - \frac{6q^2t^2}{2^{n+1}}\right).$$

Hence we show that (2) is established with $\varepsilon_2 = \frac{6q^2t^2}{2^{n+1}}$ .

Consequentially, we obtain the upper bound on advantage of adversary $\mathcal{A}$

$$\mathbf{Adv}^{prf}_{DP[\pi]}(\mathcal{A}) \leq \varepsilon_1 + \varepsilon_2 = \frac{7t^2q^2}{2^{n+1}}\ . \qquad \square$$

On the other hand, any random permutation $\pi \in \mathcal{P}(\{0,1\}^n)$ in Theorem 3 is implemented by a secure block cipher in the real field. Hence it is important to investigate the concrete security analysis for the PRP-based Double-Pipeline Iteration mode where the underlying PRP is a practical block cipher such as AES. Theorem 4 shows that the Double-Pipeline Iteration mode using a block cipher is also secure, if the underlying block cipher is secure on the view point of concrete security paradigm. Let $E_K$ be a permutation family where $K$ is randomly chosen from $\mathcal{K}$. Then a block cipher with key is regarded as a instance of the $E_K$.

**Theorem 4.** *Let $\mathcal{A}$ be any prf-adversary attacking $DP[E_K]$ with any $q$ queries such that $\frac{t^2q^2}{2^n} \leq \frac{2}{3}$ . Then there exists an prp-adversary $\mathcal{B}$ attacking $E_K$ with $2tq$ queries such that*

$$\mathbf{Adv}^{prf}_{DP[E_K]}(\mathcal{A}) \leq \mathbf{Adv}^{prp}_{E_K}(\mathcal{B}) + \frac{7t^2q^2}{2^{n+1}} .$$

***Proof.*** We specify an adversary $\mathcal{B}$ attacking $E_K$. Let $E_K$ be a permutation family with induced distibution from $\mathcal{K}$ , $\pi$ be a random permutation from $\mathcal{P} = \mathcal{P}(\{0,1\}^n)$. The adversary $\mathcal{B}$ will get an oracle for a permutation $g$ on $\{0,1\}^n$. In World 0, $g$ will be chosen from $\mathcal{P}$, that is, $g = \pi$, while in World 1, $g$ will be set to $E_K$ where $K$ is a randomly chosen key. The adversary $\mathcal{B}$ will run $\mathcal{A}$ as a subroutine. The $\mathcal{B}$ work like this:

Adversary $\mathcal{B}^g$
  Run adversary $\mathcal{A}$, replying to its oracle queries as follows
    For $i = 1, \cdots, q$ do
    When $\mathcal{A}$ makes an oracle query $x^{(i)}$
      $\mathbf{z}^{(i)} \xleftarrow{\$} DP[g](x^{(i)})$
      Return $\mathbf{z}^{(i)}$ to $\mathcal{A}$ as the answer
    Until $\mathcal{A}$ stops and outputs a bit, $b$
  Return $b$

Then by Definition 1, prp-adventage of $\mathcal{B}$ is

$$\begin{aligned}\mathbf{Adv}^{prp}_{E_K}(\mathcal{B}) &= Pr\left[\mathbf{Exp}^{prp-1}_{E_K}(\mathcal{B}) = 1\right] - Pr\left[\mathbf{Exp}^{prp-0}_{E_K}(\mathcal{B}) = 1\right] \\ &= Pr\left[\mathcal{B}^g = 1 \mid g \leftarrow E_K\right] - Pr\left[\mathcal{B}^g = 1 \mid g \leftarrow \mathcal{P}\right].\end{aligned}$$

In this case,

$$Pr\left[\mathcal{B}^g = 1 \mid g \leftarrow E_K\right] = Pr\left[\mathcal{A}^G = 1 \mid G \leftarrow DP[E_K]\right] ,$$

$$Pr\left[\mathcal{B}^g = 1 \mid g \leftarrow \mathcal{P}\right] = Pr\left[\mathcal{A}^G = 1 \mid G \leftarrow DP[\pi]\right].$$

Therefore,

$$\begin{aligned}\mathbf{Adv}^{prp}_{E_K}(\mathcal{B}) &= Pr\left[\mathbf{Exp}^{prp-1}_{E_K}(\mathcal{B}) = 1\right] - Pr\left[\mathbf{Exp}^{prp-0}_{E_K}(\mathcal{B}) = 1\right] \\ &= Pr\left[\mathcal{B}^g = 1 \mid g \leftarrow E_K\right] - \left[\mathcal{B}^g = 1 \mid g \leftarrow \mathcal{P}\right] \\ &= Pr\left[\mathcal{A}^G = 1 \mid G \leftarrow DP[E_K]\right] - Pr\left[\mathcal{A}^G = 1 \mid G \leftarrow DP[\pi]\right] \\ &= Pr\left[\mathcal{A}^G = 1 \mid G \leftarrow DP[E_K]\right] - Pr\left[\mathcal{A}^G = 1 \mid G \leftarrow \mathcal{F}\right] \\ &\quad + Pr\left[\mathcal{A}^G = 1 \mid G \leftarrow \mathcal{F}\right] - Pr\left[\mathcal{A}^G = 1 \mid G \leftarrow DP[\pi]\right] \\ &= \mathbf{Adv}^{prf}_{DP[E_K]}(\mathcal{A}) - \mathbf{Adv}^{prf}_{DP[\pi]}(\mathcal{A}) .\end{aligned}$$

By Theorem 3, we obtain that

$$\mathbf{Adv}^{prf}_{DP[E_K]}(\mathcal{A}) \leq \mathbf{Adv}^{prp}_{E_K}(\mathcal{B}) + \frac{7t^2q^2}{2^{n+1}} . \qquad \square$$

## 5 Conclusion

In this paper we have examined the soundness of the PRP-based KEFs, variant schemes of PRF-base schemes of NIST SP 800-108, on the view point of provable security framework, and proved that the variant of Double-Pipeline Iteration mode using PRPs is secure, while the variants of Counter and Feedback modes using PRPs are insecure. Moreover we have provided a concrete security bound for the variant of Double-Pipeline Iteration mode where the underlying PRP is a practical block cipher, since in practice a secure block cipher such as AES can be regarded as a PRP. As far as we know our results are the first work related to the security analysis for the KEFs within NIST SP 800-108.

**Acknowledgements:** This research was partially supported by BK21PLUS through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (Grant No. 31Z20130012918) and the IT R&D program of MSIP/KEIT[10041864, development on Spectrum Efficient Multiband WPAN System for Smart Home Networks].

## References

1. M. Bellare, J. Kilian and P. Rogaway, *The security of the cipher block chaining message authentication code*. J. Computer and System Sciences, Vol. 61, No. 3, pp. 362-399, 2000.
2. M. Bellare and P. Rogaway, *Introduction to Mordern Cryptography*, Available at *http://cseweb.ucsd.edu/ mihir/cse207/classnotes.html*.
3. FIPS 198-1, *The Keyed-Hash Message Authentication Code (HMAC)*, 2008.
4. H. Gilbert, *The security of one-block-to-many modes of operation*, FSE 2003, Springer-Verlag, LNCS 2887, pp. 376-395, 2003.
5. O. Goldreich, S, Goldwasser and S. Micali, *How to construct random functions*, J. of the ACM, Vol. 33, No. 4, pp. 210-217, 1986.
6. H. Krawczyk, *Cryptographic extraction and key derivation: The HKDF scheme*, CRYPTO 2010, Springer-Verlag, LNCS 6223, pp. 631-648, 2010.
7. M. Luby and C. Rackoff, *How to construct pseudorandom permutations and pseudorandom functions*, SIAM J. Comput., Vol. 17, pp. 373-386, 1988.
8. M. Naor and O. Reingold, *On the construction of pseudorandom permutations: Luby-Rackoff revisited*, J. Cryptology, Vol. 12, pp. 29-66, 1999.
9. NIST Special Publication 800-108, *Recommendation for Block Cipher Modes of Operation - The CMAC Mode for Authentication*, May 2005.
10. NIST Special Publication 800-56C, *Recommendation for Key Derivation through Extraction-then-Expansion*, November 2011.
11. NIST Special Publication 800-108, *Recommendation for Key Derivation Using Pseudorandom Functions*, October 2009.
12. J. Patarin, *How to construct pseudorandom and super pseudorandom permutations from one single pseudorandom function*, Advances in Cryptology - Eurocrypt'92, Springer-Verlag, LNCS 658, pp. 256-266, 1992.
13. 3rd Generation Partnership Project, *Specification of the MILENAGE algorithm set: An example algorithm set for the 3GPP authentication and key generation functions f1, f1*, f2, f3, f4, f5 and f5**, Available at *http://www.3gpp.org*.