A preliminary version of this paper appears in the Proceedings of CCS 2016. This is the full version.

Message-recovery attacks on Feistel-based Format Preserving Encryption

Mihir Bellare [1] Viet Tung Hoang[2] Stefano Tessaro[3]

August 25, 2016

Abstract

We give attacks on Feistel-based format-preserving encryption (FPE) schemes that succeed in message recovery (not merely distinguishing scheme outputs from random) when the message space is small. For one byte messages, the attacks fully recover the target message using 2^{32} examples for the FF3 NIST standard and 2^{40} examples for the FF1 NIST standard. The examples include only three messages per tweak, which is what makes the attacks non-trivial even though the total number of examples exceeds the size of the domain. The attacks are rigorously analyzed in a new definitional framework of message-recovery security. The attacks are easily put out of reach by increasing the number of Feistel rounds in the standards.

[1] Department of Computer Science & Engineering, University of California San Diego, La Jolla, California 92093, USA. Email: mihir@eng.ucsd.edu. URL: http://cseweb.ucsd.edu/~mihir/. Supported in part by NSF grant CNS-1526801, ERC Project ERCC FP7/615074 and a gift from Microsoft.

[2] Department of Computer Science, Florida State University, Tallahassee, Florida 32304, USA. Email: tvhoang@cs.fsu.edu. URL: http://www.cs.fsu.edu/~tvhoang/. Supported in part by NSF grants CNS-1423566 and CNS-1553758 (CAREER), and by the Glen and Susanne Culler Chair. Hoang's work was done while at UCSB.

[3] Department of Computer Science, University of California Santa Barbara, Santa Barbara, California 93106, USA. Email: tessaro@cs.ucsb.edu. URL: http://www.cs.ucsb.edu/~tessaro/. Supported in part by NSF grants CNS-1423566 and CNS-1553758 (CAREER), and by the Glen and Susanne Culler Chair.

Contents

1 Introduction

Format-preserving encryption (FPE) schemes based on Feistel were standardized by NIST in [7] and are in widespread use for the encryption of credit card numbers. This paper gives new attacks on these schemes that succeed in message recovery in the case that the message space is small.

FPE. An FPE scheme [1, 5] specifies a deterministic encryption function $\mathsf{F.E} : \mathsf{F.Keys} \times \mathsf{F.Twk} \times \mathsf{F.Dom} \to \mathsf{F.Dom}$ that takes a key K, a tweak T and a message X to return a ciphertext $Y = \mathsf{F.E}(K, T, X)$. There is a corresponding decryption function $\mathsf{F.D} : \mathsf{F.Keys} \times \mathsf{F.Twk} \times \mathsf{F.Dom} \to \mathsf{F.Dom}$ such that the maps $\mathsf{F.E}(K, T, \cdot), \mathsf{F.D}(K, T, \cdot)$ are permutations over $\mathsf{F.Dom}$ that are inverses of each other.

What makes an FPE scheme special —compared to a tweakable blockcipher [8]— is that the domain $\mathsf{F.Dom}$ can be arbitrary, and, most importantly, can be very small. Some examples are $\mathsf{F.Dom} = \{0,1\}^8$ (encrypt a byte so that the ciphertext is also a byte), $\mathsf{F.Dom} = \mathbb{Z}_{10}^4$ (encrypt a 4 digit PIN so that the ciphertext is also four decimal digits), $\mathsf{F.Dom} = \mathbb{Z}_{10}^{16}$ (encrypt a 16-digit credit-card number so that the result is also a 16-digit credit-card number). FPE is motivated by legacy constraints which in many systems mandate that the ciphertext replace the plaintext, and must thus have the same "format" as the plaintext.

Schemes and standardization. FPE is harder than it may look. Blockciphers like AES can encipher 128-bit messages but it isn't clear how to encipher messages of length significantly *shorter* than 128. The main paradigm for FPE has been to use a Feistel network. Feistel based FPE schemes were given in [5, 1]. (Tweaking is done by incorporating the tweak as an input to the round functions.) The growing use of FPE led to interest in standardization. Several submissions were made to NIST [4, 6, 3]. Based on these, in March 2016, NIST SP 800-38G [7] standardized two Feistel-based FPE schemes, FF1 and FF3.

In Fig. 3 we specify Feistel-based FPE in a general and parameterized way. Prior schemes, including the standards, are special cases, and our attacks apply to all of these.

Summary. This paper has three main contributions: (1) New **message recovery attacks** on Feistel-based FPE that are practical for small messages (2) A **definitional framework** for message recovery security that allows us to precisely say what our attacks accomplish and why they are interesting (3) Rigorous **analyses** establishing lower bounds on the advantages of the attacks in our framework.

For the purpose of this Introduction, we take, as illustrative example, balanced Feistel with $\mathsf{F.Dom} = \{0,1\}^{2n}$. We denote by $r \geq 2$ the number of Feistel rounds. It is $r = 10$ for FF1 and $r = 8$ for FF3. Prior attacks and our new ones are summarized in Fig. 1. Our attacks are the first to have all of the following properties: they succeed in (partial or full) recovery of the target message, not just in distinguishing outputs of the FPE from random; they have advantage as close to one as possible, rather than very small; and they succeed given a number Q of examples —an example is a tweak, ciphertext pair (T, Y) possessed by the adversary— that, for the values of r in the standards, makes the attacks feasible for small n.

There has been a misconception that attacks using a number Q of examples in excess of the size 2^{2n} of the domain are uninteresting. This is not necessarily true. It depends on the nature of the examples. Amongst the elements that make our attacks non-trivial are that they involve only a tiny number q_e of examples for any particular tweak and encryptions of the target message are provided only under a tiny number of tweaks. In this case, FPE ought to provide very good security even for a number Q of examples well in excess of the domain size. We are saying that Feistel-based FPE with the standardized number of rounds fails to do so on small message spaces.

Prior definitions for message recovery security [1] cannot capture the distinctions we make

Attack Name	Attack type	Advantage ϵ	Number of tweaks, q_t	Examples per tweak, q_e	Source
	Distinguishing	$2^{-(r-2)n}$	1	2	[9]
	Distinguishing	1/3	$2 \cdot 2^{(r-2)n}$	2	[11]
	Partial recovery (left half)	$2^{-(r-2)n}$	1	2	[1]
LHR	Partial recovery (left half)	$1 - 2/2^n$	$24(n+4) \cdot 2^{(r-3)n}$	2	Here
RHR	Partial recovery (right half)	$1 - 2/2^n$	$24(n+4) \cdot 2^{(r-2)n}$	2	Here
FMR	Full recovery (entire message)	$1 - 2/2^n$	$24(n+4) \cdot 2^{(r-2)n}$	3	Here

Figure 1: **Attack parameters and effectiveness.** This is for balanced-Feistel FPE with domain $\{0,1\}^{2n}$ ($n \geq 2$) and r rounds. We show the type of attack (distinguishing from random, or recovery, of part or all of the message), and the advantage (success probability of the attack). The number of examples used is $Q = q_t \cdot q_e$, broken down into the number q_t of tweaks involved and the number q_e of examples per tweak. The running time of all attacks is $\mathcal{O}(Q)$.

$2n$	$r = 8$ (FF3)		$r = 10$ (FF1)	
	ϵ	Q	ϵ	Q
4	1/2	2^{21}	1/2	2^{25}
8	14/16	2^{32}	14/16	2^{40}
14	63/64	2^{53}	63/64	2^{67}

Figure 2: **Attack numbers.** We show the advantage and number of examples for the FMR attack for various input lengths $2n$ and the number of rounds of the standards.

above. The purpose and value of our new definitional framework for message-recovery security is to elucidate when an attack is non-trivial. It is so when the adversary advantage under our definition is large. Beyond this, our framework allows us to capture fine-grained distinctions between attacks (for example, full versus partial plaintext recovery, known versus unknown example plaintexts, ...) allowing theorem statements about attacks that are correspondingly fine-grained and informative.

It is common to have theorems, making precise statements and giving rigorous proofs, in support of security. Such theorems give upper bounds on adversary advantage. It is less common than it should be to have similarly rigorous theorems about attacks, giving lower bounds on adversary advantage. We give such theorems for our attacks, in the model where the Feistel round functions are random. The analyses establishing this were challenging and also allow us to give rigorous and improved analyses of some prior attacks.

The second table in Fig. 2 shows, for the full message recover (FMR) attack on the standardized schemes FF3 and FF1, the advantage ϵ and number of examples Q, for different message lengths $2n$. The attack is feasible for 4-bit messages and 8-bit messages. At 14-bit messages —this corresponds roughly to four decimal digits, the subset of the digits of the credit-card number that is encrypted in many FPE-based credit-card transactions— the attacks are not practical. In all cases, the attacks are significantly faster for FF3 than for FF1. We note that the standard [7] requires messages to have length at least $\log_{\mathsf{rdx}}(100)$ where rdx, the radix, is the size of the alphabet, so with $\mathsf{rdx} = 2$ the minimum allowed even length would be $2n = 8$ bits.

The attacks can be defended against quite simply by increasing the number of rounds on small inputs. The BRS [4] submission to NIST had, in fact, specified the number of rounds as a function of the input length. The $r = 10$ rounds adopted by NIST for FF1 was based on BRS's later addendum [3]. BPS [6] had proposed $r = 8$ rounds from the start, and this was adopted for FF3 [7]. Reverting to the formula of BRS [4] would put our attacks out of reach for the message spaces they consider. Their suggestion, for FF1 on domain $\{0,1\}^m$, was to use $r = 12$ rounds if $31 \leq m \leq 128$, $r = 18$ rounds if $20 \leq m \leq 31$, $r = 24$ rounds if $14 \leq m \leq 19$, $r = 30$ rounds if $10 \leq m \leq 13$, and $r = 36$ rounds if $8 \leq m \leq 9$.

We now expand on all the above. Our starting point is definitions, meaning attack types and taxonomy, because this is crucial towards determining the effectiveness of attacks.

Types of attacks. A distinguishing attack aims to violate (tweakable) PRP security [1, 8]. The adversary has an oracle taking T, X and returning Y such that either $Y = \mathsf{F.E}(K, T, X)$ for the target key K or Y is the result of a tweak-determined random permutation on X. In this case the examples $(T_1, X_1, Y_1), \ldots, (T_Q, X_Q, Y_Q)$ are triples where Y_i is the result of the oracle on T_i, X_i. The advantage is that of determining the type of the oracle.

Distinguishing attacks have not been considered a significant threat in practice because they do not, in general, appear to cause any practical damage in envisaged applications of FPE. The concern in practice, rather, is message recovery.

BRRS [1] give the first definition of message recovery security for FPE. The adversary gets input (T^*, Y^*) where $Y^* = \mathsf{F.E}(K, T^*, X^*)$, and its goal is to recover the target message X^*. To aid in this task, it is allowed $Q - 1$ queries to an encryption oracle. The latter, given T, X returns $Y = \mathsf{F.E}(K, T, X)$. The advantage of the adversary is the probability that it wins (returns X^*) minus the probability that a simulator, on input T^*, wins (returns X^*), given $Q - 1$ queries to a test oracle. The latter, given X, returns true if $X = X^*$ and false otherwise. The intuition is that since $\mathsf{F.E}(K, \cdot, \cdot)$ is deterministic, the adversary can use its encryption oracle to test candidate plaintexts, so the simulator gets the same ability via its test oracle.

Notice that the simulator can always win with probability one when $Q \geq 2^{2n}$ is more than the size of the domain, because it can simply query all possible messages to its test oracle. Thus, *any* adversary making $Q - 1 \geq 2^{2n} - 1$ queries to its encryption oracle has *zero advantage*. Based in part on this, the conception in this area has been that an attack using a number of examples larger than the size of the domain is trivial and not interesting.

Our framework. We argue that the above conclusion is incorrect. Attacks can be interesting, non-trivial and of practical significance even when the number of examples is much more than the size of the domain. We give a new definition for message recovery security in which this and other distinctions surface.

In our framework of Section 4, an algorithm XS called a message sampler produces Q tweak-message pairs $(T_1, X_1), \ldots, (T_Q, X_Q)$, a target message X^* and *auxiliary information* a. Now let $Y_i = \mathsf{F.E}(K, T_i, X_i)$ for $1 \leq i \leq Q$, where K is the target key. The adversary $\mathcal{A}$ gets examples $(T_1, Y_1), \ldots, (T_Q, Y_Q)$, as well as a. It wins if it outputs the target message X^*. Its mr-advantage is its winning probability minus what we call the mg-advantage of XS. The latter is the maximum, over all simulators $\mathcal{S}$, of the probability that $\mathcal{S}$, given $T_1, \ldots, T_Q$ and a, returns X^* . The auxiliary information encodes partial information about the messages that the adversary may have. There are no oracles involved.

Now, there are many choices of XS for which we would expect and want the mr-advantage to be small, even for Q much larger than the domain size. The instance we consider here is that, in the list $(T_1, X_1), \ldots, (T_Q, X_Q)$, the number q_e of times any particular tweak T shows up is very small, much smaller than the size 2^{2n} of the domain, or the number of i such that $X_i = X^*$ is very

small. So if X^* is (say) random, the mr-advantage should be small. Our attacks say that, for the standardized schemes, this advantage is not small.

The problem with the BRRS definition [1] is that the simulator queries may all be under the target tweak even if the adversary makes few queries under the target tweak. Our definition models security more accurately by forcing the simulator to use exactly the same tweaks as the adversary.

PRIOR ATTACKS. Row 1 of Fig. 1 is a distinguishing attack of Patarin [9] that in $Q = 2$ examples gets a distinguishing advantage $\epsilon \approx 2^{-(r-2)n/2}$. Row 2 is a variant he gives in [11] which achieves a constant distinguishing advantage using $Q = 4 \cdot 2^{(r-2)n}$ examples.

BRSS [1] extend Patarin's ideas [9, 10, 11] to give a message recovery attack under their definition discussed above. It recovers the left half of a message with known right half. Thus the target message $X^* = (L^*, R^*)$ has a random left half L^* and an adversary-known right half R^*. Given T^* and target ciphertext $Y^* = (A^*, B^*) = \mathsf{F.E}(K, T^*, X^*)$, the adversary picks a random left-half L and queries its encryption oracle with T^*, X for $X = (L, R^*)$ to get back $Y = \mathsf{F.E}(K, T^*, X)$. It returns $(A \oplus A^* \oplus L, R^*)$. The advantage as per the BRRS definition is about $2^{-(r-2)n}$.

These attacks were known at the time of standardization but not considered significant. In the case of Patarin's attacks, this is because they are distinguishing attacks that did not appear to cause any practical damage in envisaged applications of FPE. In the case of the BRRS attack, the advantage seems too tiny to matter. For example. say $n = 4$ (one byte messages). The a priori probability of guessing the target message is 2^{-4}. The attack recovers the target message with a probability only marginally higher, namely $2^{-4}(1 + 2^{-28}) \approx 2^{-4}$ in the case $r = 10$ (FF1). The concern in practice is message recovery with high advantage.

OVERVIEW OF OUR ATTACKS. Our attacks boost the message recovery advantage to close to one. This is done by using more examples than BRRS, but, importantly, there are very few examples for any given tweak. Our LHR attack, like the one of BRRS [1], recovers the left half of the message when the right half is known. Our RHR attack recovers the right half of the message when the left is known, but using different and more novel techniques. We then put these together to get the FMR attack recovering the entire target message. The attack parameters are shown in Fig. 1, and we now discuss the attacks at a more technical level.

The LHR and RHR attacks target a sampler XS which, for two plaintexts X and X', produces $2q_t$ tweak-message pairs $(T_1, X'), (T_1, X), \ldots, (T_{q_t}, X')$, target message $X^* = X$, and some side information a about X and X'. In particular, the end goal is recovering $X^* = X$. Here, we illustrate the main ideas behind the attacks for the special case of r-round balanced Feistel with $\mathsf{F.Dom} = \{0,1\}^{2n}$.

The LHR attack assumes that X and X' share the same right half R and have different left halves $L \neq L'$. Here, $a = (L', R)$. The starting point is Patarin's observation [9, 11] that if L_r and L'_r are left halves of the encryptions of X and X' under some tweak, then $L_r \oplus L'_r \oplus L'$ is more likely to be L than any other value. This property was exploited already in the aforementioned distinguishing attacks [9, 11] and in the low-advantage recovery attack from [1]. In contrast, here we show that under many tweaks, this fact can be exploited to recover L with constant probability – namely, if $L_{i,r}$ and $L'_{i,r}$ are the left halves of the encryptions of X and X' under T_i, respectively, the attack analyzes the empirical distribution of the values $L_{i,r} \oplus L'_{i,r} \oplus L'$, and takes the most frequent value as the guess for L. Our analysis shows that $q_t = O(2^{(r-3)n})$ suffices for the guess to be correct with constant probability. While q_t is well above the domain size, the crucial point is that we only obtain two ciphertexts *per tweak* for the same two plaintexts, and this should not help for non-trivial message recovery.

In the RHR attack, the plaintexts X and X' are distinct, but do not satisfy any other relation. Also, $a = (L, R')$, and the attack recovers the right half R of X. To understand the main ideas

behind the attack, assume we are given two encryptions of X and X' such that the left halves L_r and L'_r of the ciphertexts are equal, while their right halves $R_r \neq R'_r$ differ. Then, we show that $R' \oplus R_r \oplus R'_r$ is more likely to equal R than any other value. Intuitively, the reason for this is the similarity between evaluating Feistel in the forward and backward direction, combined with the ideas from the LHR attack. However, making this precise requires more work. Also, this by itself is not useful – we have no control on whether $L_r = L_{r'}$ occurs or not. However, on average, this will be true once every (roughly) 2^n tweaks. Indeed, we show that the number of examples required for the RHR attack to succeed is indeed 2^n larger than for the LHR attack, i.e., $q_t = O(2^{(r-2)n})$.

The final FMR attack combines both attacks, and recovers X when given *three* ciphertexts per tweak of plaintexts X, X' and X^*. In conjunction with ciphertexts of X, those of X' will be used to perform the RHR attack first, and this will then allow, together with ciphertexts of X^*, performing the LHR attack.

2 Notation and standard definitions

We let ε denote the empty string. If y is a string then $|y|$ denotes its length and $y[i]$ denotes its i-th bit for $1 \leq i \leq |y|$. If X is a finite set, we let $x \xleftarrow{\$} X$ denote picking an element of X uniformly at random and assigning it to x. Algorithms may be randomized unless otherwise indicated. Running time is worst case. If A is an algorithm, we let $y \leftarrow A(x_1, \ldots; r)$ denote running A with random coins r on inputs $x_1, \ldots$ and assigning the output to y. We let $y \xleftarrow{\$} A(x_1, \ldots)$ be the result of picking r at random and letting $y \leftarrow A(x_1, \ldots; r)$. We use the code based game playing framework of [2]. By $\Pr[\mathsf{G}]$ we denote the event that the execution of game G results in the game returning true. If D is a set then $\mathrm{Perm}(\mathsf{D})$ denotes the set of all permutations on D. Let $\exp(x)$ denote e^x, where e is the base of the natural logarithm.

3 FPE and Feistel-based FPE

FPE. A format-preserving encryption (FPE) scheme F specifies a deterministic encryption algorithm $\mathsf{F.E} : \mathsf{F.Keys} \times \mathsf{F.Twk} \times \mathsf{F.Dom} \to \mathsf{F.Dom}$ and a deterministic decryption algorithm $\mathsf{F.D} : \mathsf{F.Keys} \times \mathsf{F.Twk} \times \mathsf{F.Dom} \to \mathsf{F.Dom}$. The sets $\mathsf{F.Keys}$, $\mathsf{F.Twk}$ and $\mathsf{F.Dom}$ are, respectively, the key space, the tweak space and the domain. For every key $K \in \mathsf{F.Keys}$ and tweak $T \in \mathsf{T}$, the maps $\mathsf{F.E}(K, T, \cdot)$, $\mathsf{F.D}(K, T, \cdot) \in \mathrm{Perm}(\mathsf{F.Dom})$ are permutations over $\mathsf{F.Dom}$ that are inverses of each other.

Feistel-based FPE. Feistel-based constructions represent the currently most important method to obtain FPE. The FF1 and FF2 standards [7] are both Feistel based. We now specify Feistel-based FPE in a general, parameterized way. Particular choices of the parameters allow us to talk of schemes with ideal round functions or with concrete ones, and to recover the standards.

We associate to parameters $r, M, N, \boxplus, \mathrm{PL}$ an FPE scheme $\mathsf{F} = \mathbf{Feistel}[r, M, N, \boxplus, \mathrm{PL}]$. Here $r \geq 2$ is an even integer, the number of rounds. Integers $M, N \geq 1$ define the domain of F as $\mathsf{F.Dom} = \mathbb{Z}_M \times \mathbb{Z}_N$. Let $\boxplus$ be an operation for which $(\mathbb{Z}_M, \boxplus)$ and $(\mathbb{Z}_N, \boxplus)$ are Abelian groups. We let $\boxminus$ denote the inverse operator of $\boxplus$, meaning that $(X \boxplus Y) \boxminus Y = X$ for every X and Y. $\mathrm{PL} = (\mathcal{T}, \mathcal{K}, F_1, \ldots, F_r)$ is a list. It specifies the set $\mathcal{T}$ of tweaks, meaning $\mathsf{F.Twk} = \mathcal{T}$. It specifies a set $\mathcal{K}$ of keys, so that $\mathsf{F.Keys} = \mathcal{K}$. Finally it specifies round functions $F_1, \ldots, F_r$ where $F_i : \mathcal{K} \times \mathcal{T} \times \mathbb{Z}_N \to \mathbb{Z}_M$ if i is odd, and $F_i : \mathcal{K} \times \mathcal{T} \times \mathbb{Z}_M \to \mathbb{Z}_N$ if i is even. The encryption and decryption functions of F are shown in Fig. 3.

The simplest instance is the boolean one, where $M = 2^m$ and $N = 2^n$ are powers of two. We identify $\mathbb{Z}_M, \mathbb{Z}_N$ with $\{0,1\}^m$ and $\{0,1\}^n$, respectively, and let $\boxplus = \oplus$ be bitwise xor. Classical

$\underline{\mathsf{F.E}(K, T, X)}$
$(L, R) \leftarrow X$
For $i = 1$ to r do
  If $(i \bmod 2 = 1)$ then $L \leftarrow L \boxplus F_i(K, T, R)$
  Else $R \leftarrow R \boxplus F_i(K, T, L)$
Return (L, R)

$\underline{\mathsf{F.D}(K, T, Y)}$
$(L, R) \leftarrow Y$
For $i = r$ to 1 do
  If $i \bmod 2 = 1$ then $L \leftarrow L \boxminus F_i(K, T, R)$
  Else $R \leftarrow R \boxminus F_i(K, T, L)$
Return (L, R)

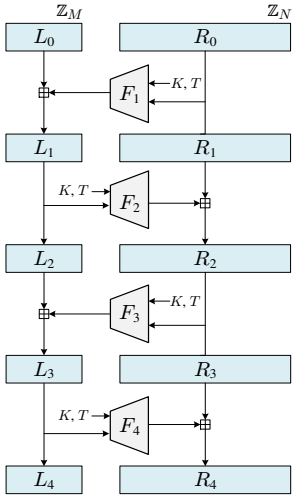

Figure 3: On the left is code for the encryption and decryption algorithms of $\mathsf{F} = \mathbf{Feistel}[r, M, N, \boxplus, \mathrm{PL}]$, where $\mathrm{PL} = (\mathcal{T}, \mathcal{K}, F_1, \ldots, F_r)$. On the right is an illustration of encryption with $r = 4$ rounds.

Feistel was, in this way, boolean. However FPE schemes sometimes operate on integers, whence the generalization. The scheme is balanced if $M = N$ and unbalanced otherwise.

We will focus on the case where the round functions are random. Proceeding formally, let $\mathbf{RF}(\mathcal{T}, r, M, N)$ denote the set of all tuples of functions $(G_1, \ldots, G_r)$ such that $G_i : \mathcal{T} \times \mathbb{Z}_N \to \mathbb{Z}_M$ if i is odd, and $G_i : \mathcal{T} \times \mathbb{Z}_M \to \mathbb{Z}_N$ if i is even. Let $\mathcal{T} = \{0,1\}^*$ and let $\mathcal{K} = \mathbf{RF}(\mathcal{T}, r, M, N)$. Then for $1 \le i \le r$ define F_i, on input K, T, X, to parse the key as $(G_1, \ldots, G_r) \leftarrow K$ and simply return $G_i(T, X)$. Now let $\mathrm{PL} = (\mathcal{T}, \mathcal{K}, F_1, \ldots, F_r)$. We write $\mathbf{Feistel}[r, M, N, \boxplus]$ to denote $\mathbf{Feistel}[r, M, N, \boxplus, \mathrm{PL}]$ for this particular choice of PL.

Schemes in the standards [7] correspond, in our framework, to particular choices of $r, M, N, \boxplus, \mathrm{PL}$. In particular they specify the round functions using AES. The analysis of our attacks, as with prior ones, is for $\mathbf{Feistel}[r, M, N, \boxplus]$, meaning round functions are truly random. However, the round functions in the standardized schemes are conjectured to be PRFs, and this means that the bounds we show on adversary advantage with random round functions translate to the standards with small differences.

For $X = (L, R) \in \mathbb{Z}_M \times \mathbb{Z}_N$, we call L and R the *left segment* and *right segment* of X, respectively. For simplicity, we assume that 0 is the zero element of the groups $(\mathbb{Z}_M, \boxplus)$ and $(\mathbb{Z}_N, \boxplus)$.

4 Message recovery framework

Here we give a new formalization of message-recovery security, defining the goal our attacks will violate.

Samplers and guessing probability. A *message sampler* is an algorithm XS that returns a tuple $((T_1, X_1), \ldots, (T_Q, X_Q), X, a)$ consisting of Q tweak-message pairs called *the example tweak-message pairs*, a message X called the *target message* and a string a called the *auxiliary information*. The *number of examples* Q is a parameter of XS that is denoted $\mathsf{XS.Q}$. We require (in our FPE context, for reasons explained below) the following *distinctness* condition: the Q pairs (T_1, X_1),

Game $\mathbf{G}^{\mathsf{mr}}_{\mathsf{F},\mathsf{XS}}(\mathcal{A})$	Game $\mathbf{G}^{\mathsf{mg}}_{\mathsf{XS}}(\mathcal{S})$
$K \leftarrow\!\!\$\ \mathsf{F.Keys}$	$((T_1, X_1), \ldots, (T_Q, X_Q), X, a) \leftarrow\!\!\$\ \mathsf{XS}$
$((T_1, X_1), \ldots, (T_Q, X_Q), X, a) \leftarrow\!\!\$\ \mathsf{XS}$	$X^* \leftarrow\!\!\$\ \mathcal{S}(T_1, \ldots, T_Q, a)$
For $i = 1, \ldots, Q$ do $Y_i \leftarrow \mathsf{F.E}(K, T_i, X_i)$	Return $(X^* = X)$
$X^* \leftarrow\!\!\$\ \mathcal{A}((T_1, Y_1), \ldots, (T_Q, Y_Q), a)$	
Return $(X^* = X)$	

Figure 4: **Games defining message-recovery security of an FPE scheme F, parameterized by a message sampler XS.**

$\ldots, (T_Q, X_Q)$ are all distinct. On the right of Fig. 4 is a *message guessing* (mg) game associated to XS and an adversary $\mathcal{S}$. Let

$$\mathbf{Adv}^{\mathsf{mg}}_{\mathsf{XS}} = \max_{\mathcal{S}} \Pr[\mathbf{G}^{\mathsf{mg}}_{\mathsf{XS}}(\mathcal{S})] \ .$$

This represents the best possible probability at guessing the target message X given the tweaks and auxiliary information. There is nothing cryptographic involved here, and the probability depends only on the message sampler.

Further parameters and terms of interest for a sampler are as follows. The *number of tweaks* of XS, denoted q_t, is the number of distinct values in the list $T_1, \ldots, T_Q$, meaning the size of the set $\{T_1, \ldots, T_Q\}$. The *number of examples per tweak*, denoted q_e, is the maximum, over all T, of the size of the set $\{\ i\ :\ T_i = T\ \}$. A tweak T is called a *target tweak* if there is some i such that $(T, X) = (T_i, X_i)$, meaning that the target message occurs with this tweak, and q^* denotes the number of target tweaks. Note that this number could be zero, one or more than one.

Message recovery security. Let F be an FPE scheme. Let XS be a message sampler such that $T_1, \ldots, T_Q \in \mathsf{F.Twk}$ and $X_1, \ldots, X_Q \in \mathsf{F.Dom}$ for any $((T_1, X_1), \ldots, (T_Q, X_Q), X, a) \in [\mathsf{XS}]$. On the left of Fig. 4 is a *message recovery* (mr) game associated to F, XS and an adversary $\mathcal{A}$. Let

$$\mathbf{Adv}^{\mathsf{mr}}_{\mathsf{F},\mathsf{XS}}(\mathcal{A}) = \Pr[\mathbf{G}^{\mathsf{mr}}_{\mathsf{F},\mathsf{XS}}(\mathcal{A})] - \mathbf{Adv}^{\mathsf{mg}}_{\mathsf{XS}} \ .$$

This measures $\mathcal{A}$'s advantage at recovering the target message given the tweaks, ciphertexts, and auxiliary information.

Discussion. The definition is a framework parameterized by the message sampler XS. An attack or a security claim can be made relative to a particular sampler or, more generally, a class of samplers. Specifying the sampler(s) allows us to precisely and formally capture attack features and draw fine-grained distinctions between attacks.

In the mr game, $X_1, \ldots, X_Q$ represent messages that the user of the target key K encrypts under tweaks $T_1, \ldots, T_Q$, respectively, so that the adversary is in possession of the tweaks, and of the ciphertexts $Y_1, \ldots, Y_Q$. The adversary is trying to recover the target message X given the tweaks and ciphertexts. The auxiliary information a represents partial information about the messages $X_1, \ldots, X_Q$ that may be known to the adversary.

A common cryptanalytic setting is a known-message attack. This would be captured in our setting by letting a be a list of all the non-target messages, meaning all X_i different from X. But our framework is more general, allowing us to capture attacks where the information the adversary has about the example messages is partial, for example their first halves. Different distributions on the data, as well as relations between the example and target message, are captured by different choices of XS.

The mg advantage captures the a priori probability of guessing the target message given the tweaks and auxiliary information. The mr advantage is the excess of the adversary's probability of winning the mr game over this mg advantage. To explain the distinctness condition on the message sampler, associate to XS the Q by Q matrix M whose (i, j)-th entry $M[i, j]$ is the boolean $((T_i, X_i) = (T_j, X_j))$. Given $Y_1, \ldots, Y_Q$, an adversary can immediately compute the entire matrix M because $\mathsf{F.E}$ is deterministic and a permutation for each fixed key and tweak. If we put no restrictions on the message sampler, we should thus give $\mathcal{S}$ the matrix M. A simpler alternative, and the one we adopted, is the distinctness condition, which effectively says that all non-diagonal entries of the matrix M are false.

In their work giving the first theoretical treatment of FPE, BRRS [1] gave a definition of message recovery security that we overviewed in Section 1. Here examples $(T_1, X_1), \ldots, (T_{Q-1}, X_{Q-1})$ are chosen by the adversary and submitted to an oracle that encrypts them under the target key and returns ciphertexts $Y_1, \ldots, Y_{Q-1}$. The adversary also knows a target tweak T^* and an encryption Y^* of the target message X^* under T^*, and wins if it finds X^*. Its advantage is relative to a simulator who gets $Q - 1$ queries to an oracle that, given X returns the boolean $(X = X^*)$. A weakness of their definition is that the simulator's test queries may use different tweaks than the ones in the examples obtained by the adversary. BRRS is concerned only with the number of queries, not their type. As a result, the definition indicates that any attack with a number of examples in excess of the domain size has zero advantage and is thus trivial and un-interesting. But many attacks using a number of examples more than the domain size are interesting and can be captured in our framework. In particular, the advantage ought to remain low if q_e is low, even if Q is high. Also, in the BRRS definition, the target message is encrypted under only one tweak. In our terminology, this means there is exactly one target tweak, $q^* = 1$. Our definition covers the target message being encrypted under multiple tweaks. Indeed, our attacks are for a situation where the number of target tweaks is large. Finally, the BRRS definition inherently captures only a chosen-plaintext attack, meaning the adversary knows the example messages in their entirety. There is no language to express the difference between attacks that know the example messages in their entirety and ones that do not, yet such a difference is important in practice.

The determinant of attack quality and non-triviality is exactly the mr-advantage as we have defined it. That an attack might be considered non-trivial, despite Q being larger than the domain, if q_e and q^* are small, is a good rule of thumb, but one must be careful in using it alone. We will discuss these parameters for our attacks but also bound the mr-advantage.

In our framework, attacks are non-adaptive, meaning examples cannot depend on prior ciphertexts. This reflects that in practice, it is such attacks that matter much more. It is much harder to mount an adaptive attack. Definitionally, the adaptive case is more complex. We can extend the mr game quite easily to this case but there are subtle issues in trying to extend the mg game that we are not sure how to address.

We view meeting our definition as a necessary but not sufficient condition for a scheme to be considered secure. That is, a feasible attack with high advantage under our definition indicates the scheme is insecure, but absence of such an attack does not necessarily mean the scheme is secure, in particular because there could be a feasible adaptive attack.

In Appendix A, we discuss the BRRS definition in more detail. Below, we'll show that the tweakable PRP notion [1, 8] implies our mr notion.

RELATION. BRRS suggest that the desirable security notion for FPE is the conventional tweakable-PRP. Recall that the tweakable-PRP advantage of an adversary $\mathcal{A}$ attacking an FPE scheme F is defined as

$$\mathbf{Adv}^{\mathsf{prp}}_{\mathsf{F}}(\mathcal{A}) = 2\Pr[\mathbf{G}^{\mathsf{prp}}_{\mathsf{F}}(\mathcal{A})] - 1,$$

Game $\mathbf{G}^{\mathsf{prp}}_{\mathsf{F}}(\mathcal{A})$	Procedure $\textsc{Enc}(T, M)$
$K \leftarrow\!\!\$\ \mathsf{F.Keys}\ ;\ b \leftarrow\!\!\$\ \{0,1\}$	If $b = 0$ then $C \leftarrow \pi_T(M)$
For $T \in \mathsf{F.Twk}$ do $\pi_T \leftarrow\!\!\$\ \mathrm{Perm}(\mathsf{F.Dom})$	Else $C \leftarrow \mathsf{F.E}(K, T, M)$
$b' \leftarrow\!\!\$\ \mathcal{A}^{\textsc{Enc}}\ ;$ Return $(b' = b)$	Return C

Figure 5: **Game defining tweakable-PRP security of an FPE scheme F.**

---

where game $\mathbf{G}^{\mathsf{prp}}_{\mathsf{F}}(\mathcal{A})$ is defined in Fig. 5. The Proposition below shows that tweakable PRP implies mr security.

Proposition 4.1 *Let* F *be an FPE scheme and* XS *be a message sampler. Then for any adversary* $\mathcal{A}$*, there's an adversary* $\mathcal{B}$ *such that*

$$\mathbf{Adv}^{\mathsf{mr}}_{\mathsf{F},\mathsf{XS}}(\mathcal{A}) \le \mathbf{Adv}^{\mathsf{prp}}_{\mathsf{F}}(\mathcal{B})\ .$$

The running time of $\mathcal{B}$ *is about the same as* $\mathcal{A}$ *plus the time to run* XS*. It makes as many queries as the number of examples that* XS *produces.*

Proof: The adversary $\mathcal{B}$ first generates $((T_1, X_1), \ldots, (T_Q, X_Q), X, a) \leftarrow\!\!\$\ \mathsf{XS}$, and then queries $C_i \leftarrow \textsc{Enc}(T_i, X_i)$ for every $i \in \{1, \ldots, Q\}$. It then runs $X^* \leftarrow\!\!\$\ \mathcal{A}((T_1, C_1), \ldots, (T_Q, C_Q), a)$. If $X^* = X$ then $\mathcal{B}$ returns 1, otherwise it returns 0. We now construct a simulator $\mathcal{S}^*$ such that

$$\mathbf{Adv}^{\mathsf{prp}}_{\mathsf{F}}(\mathcal{B}) = \Pr[\mathbf{G}^{\mathsf{mr}}_{\mathsf{F},\mathsf{XS}}(\mathcal{A})] - \Pr[\mathbf{G}^{\mathsf{mg}}_{\mathsf{XS}}(\mathcal{S}^*)],$$

and thus

$$\mathbf{Adv}^{\mathsf{prp}}_{\mathsf{F}}(\mathcal{B}) \ge \Pr[\mathbf{G}^{\mathsf{mr}}_{\mathsf{F},\mathsf{XS}}(\mathcal{A})] - \max_{\mathcal{S}}\{\Pr[\mathbf{G}^{\mathsf{mg}}_{\mathsf{XS}}(\mathcal{S})]\} = \mathbf{Adv}^{\mathsf{mr}}_{\mathsf{F},\mathsf{XS}}(\mathcal{A})\ .$$

Recall that $\mathcal{S}^*$ has inputs $T_1, \ldots, T_q, a$. We will have it create ciphertexts $Y_1, \ldots, Y_q$ and then return the result of running $\mathcal{A}$ on $(T_1, Y_1), \ldots, (T_q, Y_q), a$. We will simply have $\mathcal{S}^*$ pick $Y_1, \ldots, Y_q$ at random subject to necessary constraints, namely that the ciphertexts for the same tweaks are different. Proceeding to the details, define $\mathcal{S}^*$ as follows:

Adversary $\mathcal{S}^*(T_1, \ldots, T_q, a)$
For $i = 1, \ldots, q$ do $T \leftarrow T_i\ ;\ D_T \leftarrow \mathsf{F.Dom}$
For $i = 1, \ldots, q$ do $T \leftarrow T_i\ ;\ Y_i \leftarrow\!\!\$\ D_T\ ;\ D_T \leftarrow D_T \backslash \{Y_i\}$
$X^* \leftarrow \mathcal{A}((T_1, Y_1), \ldots, (T_q, Y_q), a)$
Return X^*

Let b be the challenge bit in game $\mathbf{G}^{\mathsf{prp}}_{\mathsf{F}}(\mathcal{B})$. Since XS satisfies the distinctness condition,

$$\Pr[\mathbf{G}^{\mathsf{mg}}_{\mathsf{XS}}(\mathcal{S}^*) = \Pr[\mathbf{G}^{\mathsf{prp}}_{\mathsf{F}}(\mathcal{B}) \Rightarrow 0 \mid b = 0],$$

whereas

$$\Pr[\mathbf{G}^{\mathsf{mr}}_{\mathsf{F},\mathsf{XS}}(\mathcal{A})] = \Pr[\mathbf{G}^{\mathsf{prp}}_{\mathsf{F}}(\mathcal{B}) \Rightarrow 1 \mid b = 1]\ .$$

Subtracting the equations above side by side,

$$\Pr[\mathbf{G}^{\mathsf{mr}}_{\mathsf{F},\mathsf{XS}}(\mathcal{A})] - \Pr[\mathbf{G}^{\mathsf{mg}}_{\mathsf{XS}}(\mathcal{S}^*) = 2\Pr[\mathbf{G}^{\mathsf{prp}}_{\mathsf{F}}(\mathcal{B})] - 1 = \mathbf{Adv}^{\mathsf{prp}}_{\mathsf{F}}(\mathcal{B})$$

as claimed. ∎

Adversary $\mathsf{LHR}((T_1, C'_1), (T_1, C_1), \ldots, (T_q, C'_q), (T_q, C_q), a)$

$X' \leftarrow a$; $L \leftarrow 0$; $(L', R) \leftarrow X'$

For $s \in \mathbb{Z}_M$ do $V_s \leftarrow 0$

For $i = 1$ to q do

  $(A, B) \leftarrow C_i$; $(A', B') \leftarrow C'_i$; $s \leftarrow A \boxminus A' \boxplus L'$; $V_s \leftarrow V_s + 1$

For $s \in \mathbb{Z}_M$ do

  If $V_s > V_L$ then $L \leftarrow s$

$X \leftarrow (L, R)$; Return X

Figure 6: **The Left-Half Recovery attack.**

5 The Left-Half Recovery attack

THE ATTACK. Our first attack is given encryptions of two samples X and X' under q tweaks $T_1, \ldots, T_q$ (for an appropriately large q), together with X', where X and X' have equal right segment, whereas their left segments differ. We do not make any assumptions on the distribution of $T_1, \ldots, T_q, X'$, but assume the left segment of X is uniform, conditioned on being distinct from the left segment of X'. Our first attack will recover X, and thus in particular its (unknown) left segment.

We formalize this using our message-recovery framework. We want to characterize under what conditions the attack works. This is done by specifying a class $\mathsf{SC1}_q$ of samplers, and then lower bounding the mr-advantage of the attack for *any* sampler in this class. We first let $\mathsf{DC1}_q$ be the class of all algorithms D that output $X' \in \mathbb{Z}_M \times \mathbb{Z}_N$ and distinct $T_1, \ldots, T_q \in \{0,1\}^*$. To any such D we associate the sampler

Sampler $\mathsf{XS}[\mathsf{D}]$

$(X', T_1, \ldots, T_q) \leftarrow\!\!\$\ \mathsf{D}$; $(L', R) \leftarrow X'$

$L \leftarrow\!\!\$\ \mathbb{Z}_N \backslash \{L'\}$; $X \leftarrow (L, R)$; $a \leftarrow X'$

Return $((T_1, X'), (T_1, X), \ldots, (T_q, X'), (T_q, X), X, a)$

The sampler $\mathsf{XS}[\mathsf{D}]$ above chooses a target message X that has the same right segment as the message X' produced by D. The number of examples is $Q = 2q$; the number of tweaks is $q_t = q$; the number of target tweaks is $q^* = q$; and the number of examples per tweak is $q_e = 2$. Since $X \neq X'$, each sampler in $\mathsf{SC1}_q$ satisfies the distinctness condition. Finally we define $\mathsf{SC1}_q = \{\mathsf{XS}[\mathsf{D}] \mid \mathsf{D} \in \mathsf{DC1}_q\}$. Note that we do not prescribe any particular behavior for D. Thus, we are considering a large class of samplers, as D ranges over $\mathsf{DC1}_q$.

Since q_e is small, we would expect and desire that adversaries have low mr-advantage, even if Q is big. Indeed, an ideal FPE scheme has this property. Our LHR attack shows that Feistel-based FPE fails to have this property. The Left-Half Recovery (LHR) attack LHR against $\mathsf{SC1}_q$ is given in Fig. 6. It can recover the left segment of X from the ciphertexts and the left segment of X'. Since our mr notion asks for full message recovery, the right segment of X is included in a, but this information is not needed for recovering the left segment of X. Theorem 5.1 below gives a lower bound on the mr advantage of LHR; this bound is illustrated in Fig. 7.

Theorem 5.1 *Let $M \geq 3, N \geq 2$ and $q \geq 1$ be integers, and let $r \geq 4$ be an even integer such that $N^{(r-2)/2} \geq 2M$. Let $\mathsf{F} = \mathbf{Feistel}[r, M, N, \boxplus]$, and let $\lambda = \left(1 - \frac{1}{M-1}\right)^2 \left(1 - \frac{1}{MN}\right)$. Then for any*

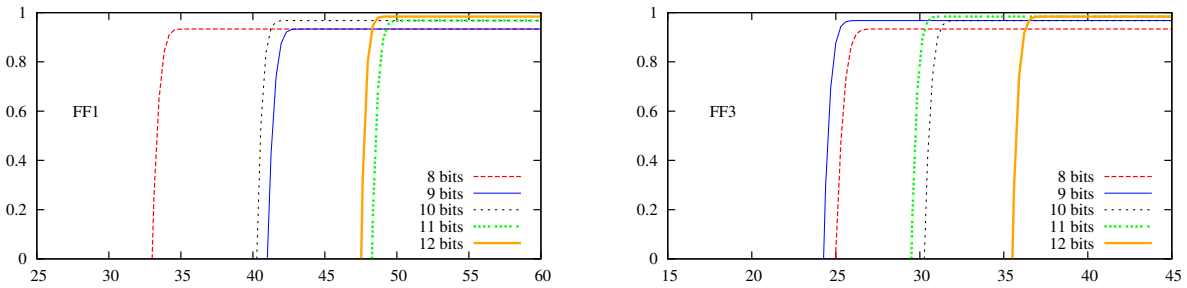

Figure 7: **The mr advantage of the Left-Half Recovery attack for binary strings of** 8–12 **bits.** The x-axis shows the log, base 2, of the number q of ciphertext pairs, and the y-axis shows $\mathbf{Adv}^{\mathsf{mr}}_{\mathbf{Feistel}[r,M,N,\boxplus],\mathsf{XS}}(\mathsf{LHR})$, for $\mathsf{XS} \in \mathsf{SC1}_q$. On the left, we use the parameters of the FF1 standard, meaning that $r = 10$, and for ℓ-bit strings, $M = 2^{\lfloor \ell/2 \rfloor}$ and $N = 2^{\lceil \ell/2 \rceil}$. On the right, we use parameters of FF3, meaning that $r = 8$, and for ℓ-bit strings, $M = 2^{\lceil \ell/2 \rceil}$ and $N = 2^{\lfloor \ell/2 \rfloor}$.

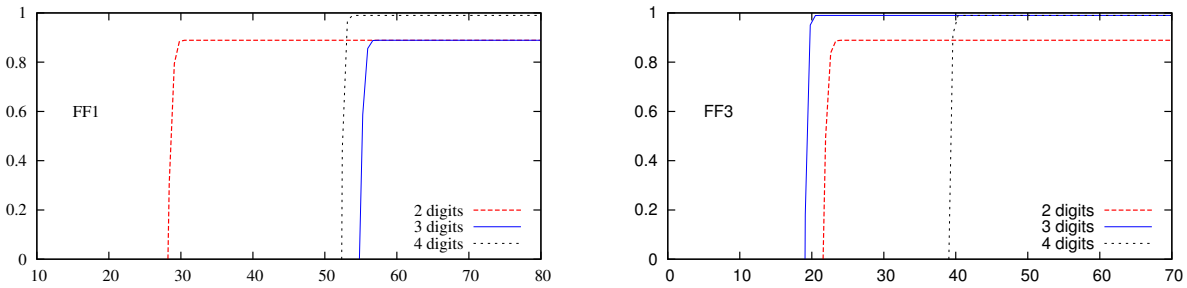

Figure 8: **The mr advantage of the Left-Half Recovery attack for decimal strings of** 2–4 **digits.** The x-axis shows the log, base 2, of the number q of ciphertext pairs, and the y-axis shows $\mathbf{Adv}^{\mathsf{mr}}_{\mathbf{Feistel}[r,M,N,\boxplus],\mathsf{XS}}(\mathsf{LHR})$, for $\mathsf{XS} \in \mathsf{SC1}_q$. On the left, we use the parameters of the FF1 standard. On the right, we use parameters of FF3.

---

sampler XS *in the class* $\mathsf{SC1}_q$,

$$\mathbf{Adv}^{\mathsf{mr}}_{\mathsf{F},\mathsf{XS}}(\mathsf{LHR}) \geq 1 - \exp\Big(\frac{-\lambda M q}{12 \cdot N^{r-2}}\Big) - M \cdot \exp\Big(\frac{-\lambda M q}{9 \cdot N^{r-2}}\Big) - \frac{1}{M-1} \ .$$

IDEAS OF THE ATTACK. The key idea of our Left-Half Recovery attack relies on the following fact, formalized and proved in Lemma 5.3, which strengthens a previous result by Patarin [9], as we explain below. Suppose that we encrypt both $X = (L, R)$ and $X' = (L', R)$ under the same tweak T. Let L_t and R_t be the left and right segments of the round-t output of X. Define L'_t and R'_t for X' likewise. Then, we show that $L_t \boxminus L'_t$ is most likely to be $L \boxminus L'$, where the probability is taken over a uniformly random choice of the key K. This is clear for $t = 1$ and $t = 2$. Indeed, $L_1 \boxminus L'_1 = L \boxminus L'$ with probability 1, and moreover, $L_2 = L_1$ and $L'_2 = L'_1$, and therefore $L_2 \boxminus L'_2 = L \boxminus L'$ is also always true. For $t = 3$, let F_3 be the third round function. Then

$$L_3 = F_3(K, T, R_2) \boxplus L_2, \text{ and } L'_3 = F_3(K, T, R'_2) \boxplus L'_2 \ .$$

If $R_2 \neq R_2'$ then $L_3 \boxminus L_3'$ is uniformly distributed over $\mathbb{Z}_M$, as L_3 and L_3' are uniform. However, if $R_2 = R_2'$, then

$$L_3 \boxminus L_3' = L_2 \boxminus L_2' = L \boxminus L' \ .$$

Therefore, at round $t = 3$, while much closer to uniform, the distribution of $L_t \boxminus L_t'$ still remains slightly biased toward the point $L \boxminus L'$. A bias remains as the round number increases, although it will decrease exponentially in t. Lemma 5.3 will concretely quantify this bias.

From the observation above, if we have ciphertexts $C = (A, B)$ and $C' = (A', B')$ of X and $X' = (L', R)$ under a single tweak T, and we know X', we can recover the left segment of X via $L = A \boxminus A' \boxplus L'$. This is exactly the message-recovery attack in [1]. However, compared to random guessing the left segment of X, this strategy only fares a little better, with advantage about $\frac{1-1/(M-1)}{N^{(r-2)/2}}$. To amplify the advantage, we need ciphertexts of X and X' under many tweaks. Hence if we have ciphertexts $(A_i, B_i) \leftarrow C_i$ and $(A_i', B_i') \leftarrow C_i'$, for $i = 1, \ldots, q$, then the Left-Half Recovery attack simply computes all values $A_i \boxminus A_i' \boxplus L'$, and output the majority value. In order to properly analyze the amplification process, we will need to develop a fine-grained understanding of the probability distribution of $L_t \boxminus L_t'$ which was not necessary in [1].

As mentioned above, we'll need to study $\Pr[L_t \boxminus L_t' = Z]$ for $Z \in \mathbb{Z}_M$. The point $Z = 0$ is an outlier; it needs a separate Lemma 5.2 below. This lemma will also be used several times in subsequent proofs for different purposes. Lemma 5.2 generalizes a result in [9] for the boolean case; the proof is in Appendix B.1.

Lemma 5.2 *Let* $\mathsf{F} = \mathbf{Feistel}[r, M, N, \boxplus]$. *Fix distinct* $X, X' \in \mathbb{Z}_M \times Z_N$, *a tweak* $T \in \mathsf{F.Twk}$, *and an even* $t \in \{2, 4, \ldots, r\}$. *Pick* $K \xleftarrow{\$} \mathsf{F.Keys}$. *Let* L_t *be the left segment of the round-*t *output of* X *under* $\mathsf{F.E}(K, T, \cdot)$. *Define* L_t' *for* X' *likewise.*

(a) If X *and* X' *have the same right segment then*

$$\tfrac{N-1}{MN-1} - \tfrac{1}{M \cdot (MN)^{(t-2)/2}} \leq \Pr[L_t = L_t'] \leq \tfrac{N-1}{MN-1}$$

(b) If X *and* X' *have different right segments then*

$$\tfrac{N-1}{MN-1} \leq \Pr[L_t = L_t'] \leq \tfrac{N-1}{MN-1} + \tfrac{1}{(MN)^{t/2}} \ .$$

Proof of Theorem 5.1. First we'll show that $\mathbf{Adv}^{\mathsf{mg}}_{\mathsf{XS}} \leq \frac{1}{M-1}$. Consider an arbitrary simulator $\mathcal{S}$. The simulator is given $X' = (L', R)$, and has to guess $X = (L, R)$, where $L \xleftarrow{\$} \mathbb{Z}_M \backslash \{L'\}$. The chance that the simulator can guess L correctly is at most $\frac{1}{M-1}$, and thus $\Pr[\mathbf{G}^{\mathsf{mg}}_{\mathsf{XS}}(\mathcal{S})] \leq \frac{1}{M-1}$. Since this bound holds for any simulator, $\mathbf{Adv}^{\mathsf{mg}}_{\mathsf{XS}} = \max_{\mathcal{S}} \Pr[\mathbf{G}^{\mathsf{mg}}_{\mathsf{XS}}(\mathcal{S})] \leq \frac{1}{M-1}$.

What's left is to show that $\Pr[\mathbf{G}^{\mathsf{mr}}_{\mathsf{F},\mathsf{XS}}(\mathsf{LHR})] \geq 1 - \exp\left(\frac{-\lambda M q}{9 \cdot N^{r-2}}\right) - M \cdot \exp\left(\frac{-\lambda M q}{12 \cdot N^{r-2}}\right)$. Recall that in the Left-Half Recovery attack, we'll iterate q times, and in the ith iteration, we'll compute a number $S_i \leftarrow A \boxminus A' \boxplus L'$, where A and A' are the left segments of the ciphertexts C_i and C_i', respectively. For each number $s \in \mathbb{Z}_M$, let $V_{i,s}$ be the Bernoulli random variable such that $V_{i,s} = 1$ if and only if $S_i = s$. The attack computes $V_s = V_{1,s} + \cdots + V_{q,s}$, finds a number z such that $V_z = \max_{s \in \mathbb{Z}_M}\{V_s\}$, and then outputs z as the left segment of the target message X. Note that for any fixed $s \in \mathbb{Z}_M$, the random variables $V_{1,s}, \ldots, V_{q,s}$ are independent and identically distributed. Let s^* be the left segment of X and $p = \frac{N}{MN-1}$. If we use an ideal FPE instead of F, then for each $s \in \mathbb{Z}_M$, $\Pr[V_{1,s} = 1]$ is exactly p. In Lemma 5.3 below, we'll show that although F is not ideal, for any $s \in \mathbb{Z}_M \backslash \{s^*\}$, $\Pr[V_{1,s} = 1] \leq p$. Yet the attack succeeds, because $\Pr[V_{1,s^*} = 1] \geq p + \Delta$, where $\Delta = \frac{1-1/(M-1)}{N^{(r-2)/2}}$. We give the proof of Lemma 5.3 further below.

Lemma 5.3 *Let* $\mathsf{F} = \mathbf{Feistel}[r, M, N, \boxplus]$. *Fix distinct* $X, X' \in \mathbb{Z}_M \times Z_N$ *of the same right segment, a tweak* $T \in \mathsf{F.Twk}$, *and an even integer* $t \in \{2, 4, \ldots, r\}$. *Pick* $K \xleftarrow{\$} \mathsf{F.Keys}$. *Let* L_t *and* L'_t *be the the left segment of the round-*t *output of* X *and* X' *under* $\mathsf{F}(K, T, \cdot)$, *respectively. Then*

(a) $\Pr[L_t \boxminus L'_t = L_0 \boxminus L'_0] \geq \frac{N}{MN-1} + \frac{1-1/(M-1)}{N^{(t-2)/2}}$.

(b) $\Pr[L_t \boxminus L'_t = Z] \leq \frac{N}{MN-1}$, *for any* $Z \in \mathbb{Z}_M \backslash \{L_0 \boxminus L'_0\}$.

The probabilities above are taken over a random sampling $K \xleftarrow{\$} \mathsf{F.Keys}$.

To analyze the advantage, our goal is to give (i) an upper bound bound for the probability that $V_s < q(p + \Delta/2)$ for every $s \in \mathbb{Z}_M \backslash \{s^*\}$, and (ii) a lower bound for the probability that $V_{s^*} > q(p + \Delta/2)$. Both (i) and (ii) are handled via Chernoff bounds.

Lemma 5.4 (Chernoff bounds) *Let* $Z_1, \ldots, Z_\ell$ *be independent Bernoulli random variables with* $\Pr[Z_1 = 1] = \cdots = \Pr[Z_\ell = 1] = p$. *Then,*

$$\begin{aligned} \Pr\Big[Z_1 + \cdots + Z_\ell \geq (1+\epsilon)\ell p\Big] &\leq \exp\Big(\frac{-\epsilon^2 \ell p}{2+\epsilon}\Big) \text{ for any } \epsilon > 0, \text{ and} \\ \Pr\Big[Z_1 + \cdots + Z_\ell \leq (1-\epsilon)\ell p\Big] &\leq \exp\Big(\frac{-\epsilon^2 \ell p}{2}\Big), \text{ for any } 0 < \epsilon < 1 \ . \end{aligned}$$

Proceeding to details, fix $s \in \mathbb{Z}_M \backslash \{s^*\}$. Let $\mu = \Pr[V_{1,s} = 1] \leq p$ and $\epsilon = \frac{\Delta}{2\mu} \geq \frac{\Delta}{2p}$. Note that $\Delta/p \leq M/N^{(r-2)/2} \leq 1/2$, and $\Delta^2/p = \lambda M/N^{r-2}$. Then

$$\frac{\epsilon^2 \mu}{2+\epsilon} = \frac{\Delta}{4/\epsilon + 2} \geq \frac{\Delta}{8p/\Delta + 2} = \frac{\Delta^2/p}{8 + 2\Delta/p} \geq \frac{\lambda M}{9 \cdot N^{r-2}} \ .$$

Since $(1+\epsilon)\mu = \mu + \Delta/2 \leq p + \Delta/2$, by the Chernoff bound,

$$\begin{aligned} \Pr[V_s \geq q(p + \Delta/2)] &\leq \Pr[V_{1,s} + \cdots + V_{q,s} \geq q(1+\epsilon)\mu] \\ &\leq \exp\Big(\frac{-\epsilon^2 \mu q}{2+\epsilon}\Big) \leq \exp\Big(\frac{-\lambda M q}{9 \cdot N^{r-2}}\Big) \ . \end{aligned} \tag{1}$$

Next, let $\mu^* = \Pr[V_{1,s^*} = 1] \geq \Delta + p$ and let $\epsilon^* = \frac{\Delta}{2(p+\Delta)}$. Then $0 < \epsilon^* < 1$. Moreover,

$$(\epsilon^*)^2 \mu^* \geq \frac{\Delta^2 q}{4(p+\Delta)} = \frac{\Delta^2/p}{4(1+\Delta/p)} \geq \frac{\Delta^2/p}{6} = \frac{\lambda M}{6 \cdot N^{r-2}} \ .$$

Since $(1-\epsilon^*)\mu^* \geq \Big(1 - \frac{\Delta}{2(p+\Delta)}\Big)(\Delta + p) = p + \Delta/2$, by the Chernoff bound,

$$\begin{aligned} \Pr[V_{s^*} \leq q(p + \Delta/2)] &\leq \Pr[V_{1,s^*} + \cdots + V_{q,s^*} \leq q(1-\epsilon^*)\mu^*] \\ &\leq \exp\Big(\frac{-(\epsilon^*)^2 \mu^* q}{2}\Big) \leq \exp\Big(\frac{-\lambda M q}{12 \cdot N^{r-2}}\Big) \ . \end{aligned} \tag{2}$$

From Equation (1) and Equation (2), the adversary LHR can correctly guess s^* with probability at least

$$1 - \Pr[V_{s^*} \le q(p+\Delta/2)] - \sum_{s \in \mathbb{Z}_M \backslash \{s^*\}} \Pr[V_s \ge q(p+\Delta/2)]$$
$$\ge 1 - \exp\Big(\tfrac{-\lambda M q}{12 \cdot N^{r-2}}\Big) - M \cdot \exp\Big(\tfrac{-\lambda M q}{9 \cdot N^{r-2}}\Big)$$

as claimed.

Proof of Lemma 5.3. Let $\mathrm{Hit}_t(Z)$ denote the event that $L_t \boxminus L'_t = Z$. Recall that we'd like to give a lower bound for $\Pr[\mathrm{Hit}_t(L_0 \boxminus L'_0)]$, and an upper bound for $\Pr[\mathrm{Hit}_t(Z)]$, for every $Z \in \mathbb{Z}_M \backslash \{L_0 \boxminus L'_0\}$. Lemma 5.2 already gives the bound for $\Pr[\mathrm{Hit}_t(0)]$. For the rest, we use the following Lemma 5.5; its proof is deferred to further below. This lemma shows that (i) $\Pr[\mathrm{Hit}_t(Z)]$ is the same for any $Z \in \mathbb{Z}_M \backslash \{0, L_0 \boxminus L'_0\}$, and (ii) the gap between $\Pr[\mathrm{Hit}_t(L_0 \boxminus L'_0)]$ and $\Pr[\mathrm{Hit}_t(Z)]$ is at least $1/N^{(r-2)t}$ for any $Z \in \mathbb{Z}_M \backslash \{0, L_0 \boxminus L'_0\}$. Combining these properties with Lemma 5.2, we have the complete picture of the distribution of $L_t \boxminus L'_t$, and thus can derive the desired bounds.

Lemma 5.5 *Let* $\mathsf{F} = \mathbf{Feistel}[r, M, N, \boxplus]$. *Fix distinct* $X, X' \in \mathbb{Z}_M \times Z_N$, $T \in \mathsf{F.Twk}$, *and an even* $t \in \{2, 4, \ldots, r\}$. *Pick* $K \leftarrow\!\!\$\ \mathsf{F.Keys}$. *Let* L_t *and* R_t *be the left and right segments of the round-*t *output of* X *under* $\mathsf{F.E}(K, T, \cdot)$, *respectively. Define* L'_t *and* R'_t *for* X' *likewise. Let* $\mathrm{Hit}_t(Z)$ *denote the event that* $L_t \boxminus L'_t = Z$; *the distribution is taken over a random sampling of* $K \leftarrow\!\!\$\ \mathsf{F.Keys}$. *Then*

1. *For any* Z, Z' *in* $\mathbb{Z}_M \backslash \{0, L_0 \boxminus L'_0\}$, *we have* $\Pr[\mathrm{Hit}_t(Z)] = \Pr[\mathrm{Hit}_t(Z')]$.
2. $\Pr[\mathrm{Hit}_t(L_0 \boxminus L'_0)] \ge \Pr[\mathrm{Hit}_t(Z)] + \frac{1}{N^{(r-2)t}}$, *for any* $Z \in \mathbb{Z}_M \backslash \{0, L_0 \boxminus L'_0\}$.

Back to the proof of Lemma 5.3, from Lemma 5.5, for any $Z \in \mathbb{Z}_M \backslash \{0, L_0 \boxminus L'_0\}$, the probability $\Pr[\mathrm{Hit}_t(Z)]$ is the same. Then for any $Z \in \mathbb{Z}_M \backslash \{0, L_0 \boxminus L'_0\}$,

$$\Pr[\mathrm{Hit}_t(Z)] = \frac{1}{M-2}\Big(1 - \Pr[\mathrm{Hit}_t(0)] - \Pr[\mathrm{Hit}_t(L_0 \boxminus L'_0)]\Big) \ .$$

Hence, once we establish the lower bound of $\Pr[\mathrm{Hit}_t(L_t \boxminus L'_t)]$, using the lower bound of $\Pr[\mathrm{Hit}_t(0)]$ as given in Lemma 5.2, the upper bound of $\Pr[\mathrm{Hit}_t(Z)]$ will automatically follow. Next, from Lemma 5.5,

$$\Pr[\mathrm{Hit}_t(L_0 \boxminus L'_0)] \ge \Pr[\mathrm{Hit}_t(Z)] + \frac{1}{N^{(r-2)t}},$$

for any $Z \in \mathbb{Z}_M \backslash \{0, L_0 \boxminus L'_0\}$, and thus

$$\begin{aligned} \Pr[\mathrm{Hit}_t(L_0 \boxminus L'_0)] &= 1 - \Pr[\mathrm{Hit}_t(0)] - \sum_{Z \in \mathbb{Z}_M \backslash \{0, L_0 \boxminus L'_0\}} \Pr[\mathrm{Hit}_t(Z)] \\ &\ge 1 - \frac{N-1}{MN-1} - (M-2) \cdot \Big(\Pr[\mathrm{Hit}_t(L_0 \boxminus L'_0)] - \frac{1}{N^{(r-2)t}}\Big). \end{aligned}$$

Hence

$$\Pr[\mathrm{Hit}_t(L_0 \boxminus L'_0)] \ge \frac{N}{MN-1} + \frac{1 - 1/(M-1)}{N^{(r-2)t}},$$

giving the claimed lower bound for $\Pr[\mathrm{Hit}_t(L_0 \boxminus L'_0)]$.

Proof of Lemma 5.5. Let F_i be the round function of F at round i, and let $G_i(\cdot, \cdot)$ be $F_i(K, \cdot, \cdot)$. We'll prove that for any Z in $\mathbb{Z}_M \backslash \{0\}$, if $t \ge 4$ then

$$\Pr[\mathrm{Hit}_t(Z)] = \tfrac{\Pr[R_{t-2} \ne R'_{t-2}]}{M} + \tfrac{\Pr[\mathrm{Hit}_{t-2}(Z)]}{N} \ . \tag{3}$$

We postpone justifying Equation (3). We now show that our claims are implied by Equation (3), via induction on t.

Proceeding to details, let's first prove the first claim. Fix Z, Z' in $\mathbb{Z}_M \backslash \{0, L_0 \boxminus L'_0\}$. First, consider the base case $t = 2$. Since $R_0 = R'_0$, we have $L_1 \boxminus L'_1 = L_0 \boxminus L'_0$, and recall that $L_2 = L_1$ and $L'_2 = L'_2$. Hence $\Pr[\mathrm{Hit}_2(Z)]$ and $\Pr[\mathrm{Hit}_2(Z')]$ are 0, and the first claim holds for the base case. Suppose that it holds for $t-2$, we'll show that it holds for t as well. From Equation (3),

$$\Pr[\mathrm{Hit}_t(Z)] = \frac{1}{M} \cdot \Pr[R_{t-2} \neq R'_{t-2}] + \frac{1}{N} \cdot \Pr[\mathrm{Hit}_{t-2}(Z)] \ .$$

Likewise,

$$\Pr[\mathrm{Hit}_t(Z')] = \frac{1}{M} \cdot \Pr[R_{t-2} \neq R'_{t-2}] + \frac{1}{N} \cdot \Pr[\mathrm{Hit}_{t-2}(Z')] \ .$$

From the induction hypothesis, $\Pr[\mathrm{Hit}_{t-2}(Z)] = \Pr[\mathrm{Hit}_{t-2}(Z')]$. Hence

$$\Pr[\mathrm{Hit}_t(Z)] = \Pr[\mathrm{Hit}_t(Z')] \ .$$

Next, we'll prove the second claim. Fix $Z \in \mathbb{Z}_M \backslash \{0, L_0 \boxminus L'_0\}$. Again, we'll prove by induction on t. First consider the base case $t = 2$. As above, $\Pr[\mathrm{Hit}_2(Z) = 0]$, while $\Pr[\mathrm{Hit}_2(L_0 \boxminus L'_0)] = 1$. Then the second claim holds for the base case. Suppose that it holds for $t-2$, we'll show that it holds for t as well. From Equation (3),

$$\Pr[\mathrm{Hit}_t(Z)] = \frac{1}{M} \cdot \Pr[R_{t-2} \neq R'_{t-2}] + \frac{1}{N} \cdot \Pr[\mathrm{Hit}_{t-2}(Z)],$$

whereas

$$\Pr[\mathrm{Hit}_t(L_0 \boxminus L'_0)] = \frac{1}{M} \cdot \Pr[R_{t-2} \neq R'_{t-2}] + \frac{1}{N} \cdot \Pr[\mathrm{Hit}_{t-2}(L_0 \boxminus L'_0)],$$

From the induction hypothesis,

$$\Pr[\mathrm{Hit}_{t-2}(L_0 \boxminus L'_0)] \geq \Pr[\mathrm{Hit}_{t-2}(Z)] + \frac{1}{N^{(r-2)(t-2)}} \ .$$

Hence the second claim also holds for t.

We now prove Equation (3). Fix $Z \in \mathbb{Z}_M \backslash \{0\}$. Note that

$$\begin{aligned} L_t &= L_{t-1} = G_{t-1}(T, R_{t-2}) \boxplus L_{t-2}, \text{ and} \\ L'_t &= L'_{t-1} = G_{t-1}(T, R'_{t-2}) \boxplus L'_{t-2} \ . \end{aligned}$$

On the one hand, since G_{t-1} is a truly random function

$$\Pr[\mathrm{Hit}_t(Z) \wedge (R_{t-2} \neq R'_{t-2})] = \frac{1}{M} \cdot \Pr[(R_{t-2} \neq R'_{t-2})]. \tag{4}$$

On the other hand, if $R_{t-2} = R'_{t-2}$ then $L_t \boxminus L'_t = L_{t-2} \boxminus L'_{t-2}$, and thus

$$\Pr[\mathrm{Hit}_t(Z) \wedge (R_{t-2} = R'_{t-2})] = \Pr[\mathrm{Hit}_{t-2}(Z) \wedge (R_{t-2} = R'_{t-2})]. \tag{5}$$

If $L_{t-2} \boxminus L'_{t-2} = Z$ then $L_{t-3} \neq L'_{t-3}$, because $L_{t-3} = L_{t-2}$ and $L'_{t-3} = L'_{t-2}$, and thus

$$\Pr[R_{t-2} = R'_{t-2} \mid \mathrm{Hit}_{t-2}(Z)] = \frac{1}{N},$$

because $R_{t-2} = G_{t-2}(T, L_{t-3}) \boxplus R_{t-3}$, $R'_{t-2} = G_{t-2}(T, L'_{t-3}) \boxplus R'_{t-3}$, and G_{t-2} is independent of L_{t-3} and L'_{t-3}. Hence

$$\Pr[\mathrm{Hit}_{t-2}(Z) \wedge (R_{t-2} = R'_{t-2})] = \frac{1}{N} \cdot \Pr[\mathrm{Hit}_{t-2}(Z)] \ . \tag{6}$$

Combining Equations (4), (5), and (6) yields Equation (3).

Comparison with prior attacks. Our attack is inspired by previous distinguishing attacks by Patarin [9, 10, 11] for the case where $M = N = 2^n$, and $\boxplus$ is the xor operator. In the first attack [9], given ciphertexts $C = (A, B)$ and $C' = (A', B')$ of two known messages $X = (L, R)$ and $X' = (L', R)$, the distinguisher outputs 1 (meaning the ciphertexts are indeed encrypted via $\mathbf{Feistel}[r, 2^n, 2^n, \oplus]$) if $A \oplus A' = L \oplus L'$, and outputs 0 (meaning the ciphertexts are encrypted via an ideal FPE) otherwise. This attack wins with advantage about $\Delta = \frac{1-1/(2^n-1)}{2^{(r-2)n/2}}$. The later attacks [10, 11] improved the advantage to constant by having ciphertexts of X and X' under many tweaks.[1] His analyses (see below for a detailed discussion) suggest $\Theta(2^{(r-2)n})$ tweaks are sufficient to distinguish with constant advantage.

Compared with Patarin's attack, our Left-Half Recovery attack is better in every front: (i) it can recover the left segment of the target message, while Patarin's attack only distinguishes the ciphertexts from random strings, (ii) it handles any domain $\mathbb{Z}_N \times \mathbb{Z}_N$ and any operator $\boxplus$, (iii) our analysis shows that $O(n \cdot 2^{(r-3)n})$ ciphertexts are sufficient, whereas Patarin's only showed the attack succeeds (in achieving a weaker goal) with a larger number of ciphertext, namely $\Theta(2^{(r-2)n})$.

To justify our comparison, we give a concise description of the most refined of Patarin's attacks [11], and sketch an analysis of the resulting advantage following Patarin's approach. (The original paper does not spell out many of these details, thus some of the following is our own interpretation.) In this distinguishing attack, one is given $((T_1, C_1, C'_1), \ldots, (T_q, C_q, C'_q))$. In the real game, C_i and C'_i are pairs ciphertexts under tweak T_i of the known messages $X = (L, R)$ and $X' = (L', R)$, respectively. In the ideal game, C_i and C'_i are uniformly chosen from $\{0,1\}^{2n}$ subject to the constraint that $C_i \neq C'_i$. Let $V_i = 1$ if $A_i \oplus A'_i = L \oplus L'$, and $V_i = 0$ otherwise, where A_i and A'_i are left segments of C_i and C'_i in the real game, respectively. Define U_i for the ideal game likewise. Let $p = \frac{2^n}{2^{2n}-1}$ and $\Delta = \frac{1-1/(2^n-1)}{2^{(r-2)n}}$. Patarin shows that $V_1, \ldots, V_q$ are independent and identically distributed Bernoulli random variables, with $\Pr[V_1] \geq p + \Delta$. Moreover, $U_1, \ldots, U_q$ are independent and identically distributed Bernoulli random variables, with $\Pr[U_1] = p$.

Let $V = V_1 + \cdots V_q$ and $U = U_1 + \cdots + U_q$. Patarin suggests that q should be picked so that $\mathbf{E}[V] - \mathbf{E}[U] \geq \sqrt{2} \cdot \big(\sqrt{\mathbf{Var}[U]} + \sqrt{\mathbf{Var}[V]}\big)$, meaning $q \approx 2 \cdot 2^{(r-2)n}$ (the additional factor $\sqrt{2}$ was not present in the original paper, but it makes calculations somewhat easier). The distinguisher receives $(T_1, C_1, C'_1), \ldots, (T_q, C_q, C'_q)$ and lets $Z_i = 1$ if $A_i \oplus A'_i = L \oplus L'$, and $Z_i = 0$ otherwise, where A_i and A'_i are left segments of C_i and C'_i respectively. It outputs 1 if $Z_1 + \cdots + Z_q \geq \mathbf{E}[V] - \sqrt{2} \cdot \sqrt{\mathbf{Var}[V]}$, and outputs 0 otherwise. In the real game, by Chebyshev's inequality, the chance that the distinguisher outputs 1 is at least

$$1 - \Pr\Big[V < \mathbf{E}[V] - \sqrt{2} \cdot \sqrt{\mathbf{Var}[V]}\Big] \geq 1 - \frac{1}{1+2} = \frac{2}{3} \ .$$

In the ideal game, the chance it outputs 1 is at most

$$\Pr\Big[U \geq \mathbf{E}[U] + \sqrt{2} \cdot \sqrt{\mathbf{Var}[U]}\Big] \leq \frac{1}{1+2} = \frac{1}{3} \ .$$

[1] Note that tweakable block ciphers were introduced by LRW [8] after Patarin's work, and Patarin's wording was of attacking "independent permutations." His attacks are however easily translated to tweakable block ciphers.

Hence, the distinguisher wins with advantage 1/3. We note that the attack complexity can in fact be reduced by using a better concentration bound (like Chernoff) to match what achieved by our attacks – however, we recall the reader that we target message recovery.

In the same work, Patarin also suggests an improved distinguishing attack, where $\ell > 2$ messages per tweak are queried. Concretely, the distinguisher picks ℓ distinct messages $X_1, \ldots, X_\ell$ and q tweaks $T_1, \ldots, T_q$. It then asks to get the corresponding ciphertexts $(C_{1,1}, \ldots, C_{1,\ell}), \ldots, (C_{q,1}, \ldots, C_{q,\ell})$ for each message-tweak pair. Let L_i be the left segment of X_i, and $A_{s,i}$ be the left segments of $C_{s,i}$. In the real game, for $s \in \{1, \ldots, q\}$ and $i, j \in \{1, \ldots, \ell\}$ such that $i < j$ and X_i and X_j have identical right segments, let $V_{s,i,j} = 1$ if $A_{s,i} \oplus A_{s,j} = L_i \oplus L_j$, and let $V_{s,i,j} = 0$ otherwise. Now, the random variables $V_{s,i,j}$ are dependent Bernoulli random variables, as multiple queries are made on the same tweak. However, Patarin conjectures that there is sufficient independence to apply the above Chebyshev argument while at the same time choosing ℓ to be sufficiently large, up to $\ell = \Theta(2^{2n})$. This would allow for (at best) $\Theta(2^{3n})$ possible pairs i, j such that X_i and X_j have identical right segment, giving us $\Theta(q2^{3n})$ random variables $V_{s,i,j}$. Under Patarin's conjecture, we can choose $q = \Theta(2^{(r-5)n})$ to have more than $2^{(r-2)n}$ variables, and apply the above argument.

This would however also result in $\Theta(2^{(r-3)n})$ ciphertexts, as in our attack, at the cost of an unproved conjecture, and still for the simpler goal of distinguishing.[2]

6 The Right-Half Recovery attack

In the Left-Half Recovery attack, the target message X and the known message X' must have the same right segment, and the attack recovers the left segment of X. In contrast, in the Right-Half Recovery attack, we have no requirement on the relationship between X and X', and the attack will recover the right segment of X.

THE ATTACK. Fix an integer $q \geq 1$. Let $\mathsf{DC2}_q$ be the class of all algorithms D that output $X' \in \mathbb{Z}_M \times \mathbb{Z}_N$ and distinct $T_1, \ldots, T_q \in \{0,1\}^*$. Let $\mathsf{SC2} = \{\mathsf{XS}[\mathsf{D}] \mid \mathsf{D} \in \mathsf{DC2}\}$, where each sampler $\mathsf{XS}[\mathsf{D}]$ in $\mathsf{SC2}_q$ behaves as follows.

Sampler $\mathsf{XS}[\mathsf{D}]$
$(X', T_1, \ldots, T_q) \leftarrow\!\!\$\ \mathsf{D}$; $(L', R') \leftarrow X'$
$(L, R) \leftarrow X \leftarrow\!\!\$\ (\mathbb{Z}_M \times \mathbb{Z}_N) \backslash \{X'\}$; $a \leftarrow (L, R')$
Return $((T_1, X'), (T_1, X), \ldots, (T_q, X'), (T_q, X), X, a)$

Here the sampler $\mathsf{XS}[\mathsf{D}]$ picks a target X that is different from the message X' produced by D. The number of examples $Q = 2q$; the number of tweaks is $q_t = q$; the number of target tweaks is $q^* = q$; and the number of examples per tweak is $q_e = 2$. Since $X \neq X'$, each sampler in $\mathsf{SC2}_q$ satisfies the distinctness condition. The Right-Half Recovery attack RHR against $\mathsf{SC2}_q$ is shown in Fig. 9. Since q_e is small, we would expect and desire that adversaries have low mr-advantage, even if Q is big. Indeed, an ideal FPE scheme has this property. Our RHR attack shows that Feistel-based FPE fails to have this property. It can recover the right segment of X from the ciphertexts and the right segment R' of X'; the left segment of X' is not needed. Since our mr notion asks for full message recovery, the auxiliary information contains the left segment L of X, but this information is not needed for recovering the right segment of X. Theorem 6.1 below gives a lower bound on the mr advantage of RHR; this bound is illustrated in Fig. 10.

[2]We note that Patarin appears to claim a lower attack complexity, but this seems to be due to a small error assuming that all pairs i, j give two inputs X_i and X_j with equal right segment for $\ell = \Theta(2^{2n})$, which is easily seen not to be possible.

Adversary $\mathsf{RHR}((T_1, C'_1), (T_1, C_1), \ldots, (T_q, C'_q), (T_q, C_q), a)$

$(L, R') \leftarrow a$; $R \leftarrow 0$; $\ell \leftarrow 0$; $p \leftarrow \frac{1}{N-1}$; $\Delta \leftarrow \frac{1-1/(N-1)}{M^{(r-2)/2}}$

For $s \in \mathbb{Z}_N$ do $V_s \leftarrow 0$

For $i = 1$ to q do

  $(A, B) \leftarrow C_i$; $(A', B') \leftarrow C'_i$

  If $A = A'$ then $s \leftarrow B \boxminus B' \boxplus R'$; $\ell \leftarrow \ell + 1$; $V_s \leftarrow V_s + 1$

For $s \in \mathbb{Z}_N$ do

  If $V_s > V_R$ then $R \leftarrow s$

If $V_R \le \ell(p + \Delta/2)$ then $R \leftarrow R'$

$X \leftarrow (L, R)$; Return X

Figure 9: **The Right-Half Recovery attack.**

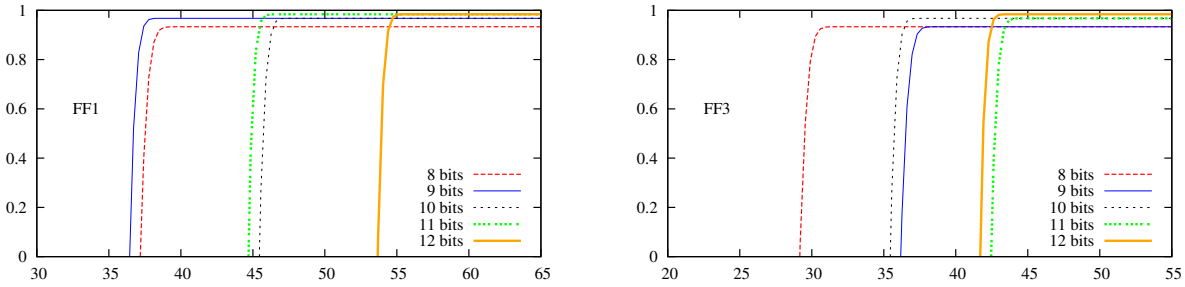

Figure 10: **The mr advantage of the Right-Half Recovery attack for binary strings of 8–12 bits.** The x-axis shows the log, base 2, of the number q of ciphertext pairs, and the y-axis shows $\mathbf{Adv}^{\mathsf{mr}}_{\mathbf{Feistel}[r,M,N,\boxplus],\mathsf{XS}}(\mathsf{RHR})$, for $\mathsf{XS} \in \mathsf{SC2}_q$. On the left, we use the parameters of the FF1 standard, meaning that $r = 10$, and for ℓ-bit strings, $M = 2^{\lfloor \ell/2 \rfloor}$ and $N = 2^{\lceil \ell/2 \rceil}$. On the right, we use parameters of FF3, meaning that $r = 8$, and for ℓ-bit strings, $M = 2^{\lceil \ell/2 \rceil}$ and $N = 2^{\lfloor \ell/2 \rfloor}$.

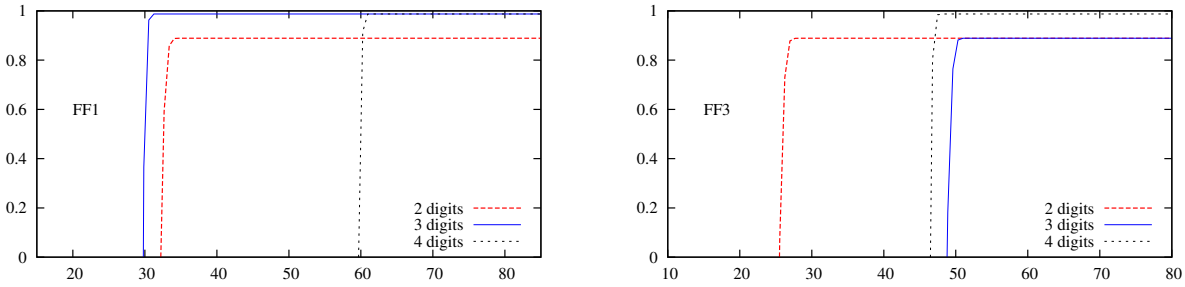

Figure 11: **The mr advantage of the Right-Half Recovery attack for decimal strings of 2–4 digits.** The x-axis shows the log, base 2, of the number q of ciphertext pairs, and the y-axis shows $\mathbf{Adv}^{\mathsf{mr}}_{\mathbf{Feistel}[r,M,N,\boxplus],\mathsf{XS}}(\mathsf{RHR})$, for $\mathsf{XS} \in \mathsf{SC2}_q$. On the left, we use the parameters of the FF1 standard. On the right, we use parameters of FF3.

Theorem 6.1 *Let $M \ge 2, N \ge 3$ and $q \ge 1$ be integers, and let $r \ge 6$ be an even integer such*

that $M^{(r-2)/2} \geq 2N$. *Let* $\lambda = \left(1-\frac{1}{N-1}\right)^2\left(1-\frac{2}{N}\right)\left(1-\frac{1}{36M^2}\right)$ *and* $\mathsf{F} = \mathbf{Feistel}[r, M, N, \boxplus]$. *Let* $\mathsf{F} = \mathbf{Feistel}[r, M, N, \boxplus]$ *and* $\lambda = \left(1-\frac{1}{N-1}\right)^2\left(1-\frac{2}{N}\right)\left(1-\frac{1}{36M^2}\right)$. *Then for any sampler* $\mathsf{XS} \in \mathsf{SC2}_q$,

$$\mathbf{Adv}^{\mathsf{mr}}_{\mathsf{F},\mathsf{XS}}(\mathsf{RHR}) \geq 1 - N \cdot \exp\Big(\frac{-\lambda qN}{9 \cdot M^{r-1}}\Big) - \exp\Big(\frac{-\lambda qN}{12 \cdot M^{r-1}}\Big) - \frac{1}{N-1} \ .$$

IDEAS OF THE ATTACK. The key idea of our Right-Half Recovery attack is based on the observation specified and proved by Lemma 6.2 below. Now, instead of requiring the known message X' and the target message X to have the same right segment, we only consider ciphertexts C and C' of $X = (L, R)$ and $X' = (L', R')$ such that C and C' have the same left segment. (Of course on average, if one encrypts X and X' under q tweaks then we only have q/M such pairs of ciphertexts.) For those ciphertexts $C = (A, B)$ and $C' = (A, B')$, the value of $B \boxminus B'$ is most likely to be $R \boxminus R'$. In some sense, this observation is the dual of the idea in Lemma 5.3 for the Left-Half Recovery. To give an intuitive (but not quite correct) explanation for this duality, note that in the boolean case, the decryption of $\mathsf{F}_1 = \mathbf{Feistel}[r, M, N, \oplus]$ is the encryption of $\mathsf{F}_2 = \mathbf{Feistel}[r, N, M, \oplus]$, but pre- and post-processed by a rotation. Then, one can imagine that, the process of encrypting X and X' via F_1 to get C and C' is effectively "encrypting" C and C' via F_2 (with additional pre- and post-processing) to get X and X'.

Then, if we have ciphertexts C_i and C'_i of a target message X and a known message $X' \leftarrow (L', R')$ under several tweaks, then we can recover the right segment of X as follows. First, keep only pairs (C_i, C'_i) such that C_i and C'_i agree on their left segments, and suppose that there are ℓ such pairs. Then, compute $P_i \leftarrow B_i \boxminus B'_i \boxplus R'$ for all such pairs, where $C_i = (A_i, B_i)$ and $C'_i = (A_i, B'_i)$. Let P be the most frequent value of those P_i. Let $p = \frac{1}{N-1}$ and $\Delta = \frac{1-1/(N-1)}{M^{(r-2)/2}}$. If X and X' agree on their right segments then expectedly, P appears ℓp times. In contrast, if X and X' differ in their right segments then P is most likely to be the right segment of X, and expectedly, it appears at least $\ell(p + \Delta)$ times. Hence if P appears at most $(p + \Delta/2)\ell$ times then we'll output R', otherwise we'll output P.

PROOF OF THEOREM 6.1. First we'll show that $\mathbf{Adv}^{\mathsf{mg}}_{\mathsf{XS}} \leq \frac{1}{N-1}$. Consider an arbitrary simulator $\mathcal{S}$. The simulator is given the right segment R' of X' and the left segment L of $X \xleftarrow{\$} (\mathbb{Z}_M \times \mathbb{Z}_N)\backslash\{X'\}$, and has to guess the right segment of X. We'll give entire X' to the simulator instead of just the right segment; it only improves the simulator's advantage. If L is also the left segment of X', then the right segment R of X is uniformly distributed over $\mathbb{Z}_N\backslash\{R'\}$, and one can guess R with probability at most $1/(N-1)$. If L is not the left segment of X' then R is uniformly distributed over $\mathbb{Z}_N$, and one can guess it with probability at most $1/N \leq 1/(N-1)$. Hence $\Pr[\mathbf{G}^{\mathsf{mg}}_{\mathsf{XS}}(\mathcal{S})] \leq \frac{1}{N-1}$. Since this bound holds for any simulator, $\mathbf{Adv}^{\mathsf{mg}}_{\mathsf{XS}} = \max_{\mathcal{S}} \Pr[\mathbf{G}^{\mathsf{mg}}_{\mathsf{XS}}(\mathcal{S})] \leq \frac{1}{N-1}$.

What's left is to show that $\Pr[\mathbf{G}^{\mathsf{mr}}_{\mathsf{F},\mathsf{XS}}(\mathsf{RHR})] \geq 1 - N \cdot \exp\Big(\frac{-\lambda qN}{9 \cdot M^{r-1}}\Big) - \exp\Big(\frac{-\lambda qN}{12 \cdot M^{r-1}}\Big)$. Recall that in the Right-Half Recovery attack, we only keep pairs (C_i, C'_i) such that C_i and C' agree on their left segments. The number of such pairs is a Binomial random variable, but for now, suppose that there are ℓ such pairs. By reindexing, let $(C_1, C'_1), \ldots, (C_\ell, C'_\ell)$ be the pairs of ciphertexts that we keep. In the attack, we iterate ℓ times, and in the i-th iteration, we'll compute $S_i \leftarrow B_i \boxminus B'_i \boxplus R'$, where $(A_i, B_i) \leftarrow C_i$ and $(A_i, B'_i) \leftarrow C'_i$. For each $s \in \mathbb{Z}_N$, let $V_{i,s}$ be the Bernoulli random variable such that $V_{i,s} = 1$ if and only if $S_i = s$. Let $p = \frac{1}{N-1}$ and $\Delta = \frac{1-1/(N-1)}{M^{(r-2)/2}}$. The attack computes $V_s = V_{1,s} + \cdots + V_{\ell,s}$, and finds $P \in \mathbb{Z}_N$ such that $V_P = \max_{s \in \mathbb{Z}_N}\{V_s\}$. If $V_P \leq \ell(p + \Delta/2)$ then we output R' as the right segment of the target message X. Otherwise, we'll output P.

Our goal is to bound the probability of $\Pr[V_s \geq \ell(p+\Delta/2)]$ for every $s \in \mathbb{Z}_N$. This can be done via Chernoff bounds, if we know the distribution of each $V_{i,s}$. For the unlikely case that X and X' have the same right segment (meaning that $R \boxminus R' = 0$), Lemma 6.2 below shows that $\Pr[V_{i,s} = 1] = p$ for every $s \in \mathbb{Z}_N \backslash \{0\}$. Intuitively, since X and X' are different, for each given tweak, they must have different ciphertexts. Since we consider only C_i and C'_i of the same left segment, they must differ in the right segment. So $B_i \boxminus B'_i$ can't be 0 in this case, but it's equally likely to be any other value in $\mathbb{Z}_N$. In this case, we're already given the right segment R of X via R'; the purpose of the ciphertexts is to help us realize that we're lucky. For the "usual" case that X and X' have different right segments, Lemma 6.2 shows that for any $s \in \mathbb{Z}_N \backslash \{R \boxminus R'\}$, the probability $\Pr[V_{i,s} = 1]$ is still bounded by p. However, for $s^* = R \boxminus R'$, the probability $\Pr[V_{i,s^*} = 1]$ jumps beyond $p + \Delta$, making the attack possible. We postpone the proof of Lemma 6.2.

Lemma 6.2 *Let* $\mathsf{F} = \mathbf{Feistel}[r, M, N, \boxplus]$. *Fix distinct* $X, X' \in \mathbb{Z}_{M,N}$, $Z \in \mathbb{Z}_N \backslash \{0\}$, $T \in \mathsf{F.Twk}$, *and an even integer* $t \in \{4, 5, \ldots, r\}$. *Pick* $K \xleftarrow{\$} \mathsf{F.Keys}$. *Let* L_t *and* R_t *denote the left and right segment of the round-*t *output of* X *under* $\mathsf{F.E}(K, T, \cdot)$. *Define* L'_t *and* R'_t *for* X' *likewise.*

(a) If $R_0 = R'_0$ *then*

$$\Pr[R_t \boxminus R'_t = Z \mid L_t = L'_t] = \frac{1}{N-1} \ .$$

(b) For $R_0 \neq R'_0$,

$$\begin{aligned} \Pr[R_t \boxminus R'_t = Z \mid L_t = L'_t] &\leq \frac{1}{N-1} \text{ if } Z \neq R_0 \boxminus R'_0, \text{ and} \\ \Pr[R_t \boxminus R'_t = Z \mid L_t = L'_t] &\geq \frac{1}{N-1} + \frac{1 - 1/(N-1)}{M^{(t-2)/2}} \text{ otherwise} \ . \end{aligned}$$

As explained above, we consider two cases for whether X and X' agree on their right segments.

Case 1: X and X' differ in their right segments. From Lemma 6.2, for any $s \in Z_N \backslash \{s^*\}$, the random variables $V_{1,s}, \ldots, V_{q,s}$ are independent and identically distributed, with $\Pr[V_{i,s} = 1] \leq p$. Likewise, $V_{1,s^*}, \ldots, V_{q,s^*}$ are independent and identically distributed, with $\Pr[V_{1,s} = 1] \geq p+\Delta$. The chance that the adversary RHR can correctly guess s^* is at least the probability that $V_s < \ell(p+\Delta/2)$ for every $s \in \mathbb{Z}_N \backslash \{s^*\}$, and $V_{s^*} > \ell(p + \Delta/2)$. To bound these probabilities, we again use Chernoff bounds.

Fix $s \in \mathbb{Z}_M \backslash \{s^*\}$. Let $\mu = \Pr[V_{1,s} = 1] \leq p$, $\epsilon = \frac{\Delta}{2\mu} \geq \frac{\Delta}{2p}$, and $z = \Delta^2/p$. Note that $\Delta/p \leq N/M^{(r-2)/2} \leq 1/2$. Then

$$\frac{\epsilon^2 \mu}{2+\epsilon} = \frac{\Delta}{4/\epsilon + 2} \geq \frac{\Delta}{8p/\Delta + 2} = \frac{\Delta^2/p}{8 + 2\Delta/p} \geq \frac{\Delta^2/p}{9} = \frac{z}{9} \ .$$

Since $(1+\epsilon)\mu = \mu + \Delta/2 \leq p + \Delta/2$, using Chernoff bounds,

$$\begin{aligned} \Pr[V_s \geq \ell(p + z \cdot \Delta)] &\leq \Pr[V_{1,s} + \cdots + V_{\ell,s} \geq \ell(1+\epsilon)\mu] \\ &\leq \exp\Big(\frac{-\epsilon^2 \mu \ell}{2+\epsilon}\Big) \leq e^{-z\ell/9} \ . \end{aligned}$$

Next, let $\mu^* = \Pr[V_{1,s^*} = 1] \geq \Delta + p$ and let $\epsilon^* = \frac{\Delta}{2(p+\Delta)}$. Then $0 < \epsilon^* < 1$. Moreover,

$$(\epsilon^*)^2\mu^* \geq \frac{\Delta^2}{4(p+\Delta)} = \frac{\Delta^2/p}{4(1+\Delta/p)} \geq \frac{\Delta^2/p}{6} = \frac{z}{6} \ .$$

Since $(1-\epsilon^*)\mu^* \geq \Big(1 - \frac{\Delta}{2(p+\Delta)}\Big)(\Delta + p) = p + \Delta/2$, using Chernoff bounds,

$$\begin{aligned} \Pr[V_{s^*} \leq \ell(p+\Delta/2)] &\leq \Pr[V_{1,s^*} + \cdots + V_{\ell,s^*} \leq \ell(1-\epsilon^*)\mu^*] \\ &\leq \exp\Big(\frac{-(\epsilon^*)^2\mu^*\ell}{2}\Big) \leq e^{-z\ell/12} \ . \end{aligned}$$

Now, given q pairs of ciphertexts, the number of pairs (C,C') among such that C and C' agree on their left segments is not a constant ℓ, but a random variable U. From Lemma 5.2, U is a Binomial random variable $B(q,\theta)$, with $\theta \geq \frac{N-1}{MN-1} \geq \frac{1-1/(N-1)}{M}$. Hence the adversary RHR can correctly guess s^* with probability at least

$$\begin{aligned} & 1 - \sum_{\ell=0}^{q} \Pr[U=\ell] \cdot \Big(\Pr[V_{s^*} \leq \ell(p+\Delta/2)] + \sum_{s \neq s^*} \Pr[V_s \geq \ell(p+\Delta/2)]\Big) \\ \geq \ & 1 - \sum_{\ell=0}^{q} \binom{q}{\ell} \theta^\ell (1-\theta)^{q-\ell} \Big(e^{-z\ell/12} + N \cdot e^{-z\ell/9}\Big) \\ = \ & 1 - \Big(\theta e^{-z/12} + 1 - \theta\Big)^q - N \cdot \Big(\theta e^{-z/9} + 1 - \theta\Big)^q \ . \end{aligned} \tag{7}$$

We first bound the term $\Big(\theta e^{-z/12} + 1 - \theta\Big)^q$ in Equation (7). From the fact that $(1-x)^q \leq e^{-qx}$ for every $0 < x < 1$,

$$\Big(\theta e^{-z/12} + 1 - \theta\Big)^q \leq \exp\Big(-q\theta(1 - e^{-z/12})\Big) \ . \tag{8}$$

Next, from the hypothesis that $r \geq 6$ and $M^{(r-2)/2} \geq 2N$,

$$z \leq \frac{1}{M^{r-2}/N} \leq \frac{1}{2 \cdot M^{(r-2)/2}} \leq \frac{1}{2M^2} \ .$$

Therefore, by using the fact that $1 - e^{-x} \geq x - x^2/2$ for every $0 < x < 1$,

$$(1 - e^{-z/12}) \geq \frac{z}{12}\Big(1 - \frac{z}{24}\Big) \geq \frac{z}{12}\Big(1 - \frac{1}{36M^2}\Big) \ .$$

Then

$$\begin{aligned} \theta(1 - e^{-z/12}) \geq \frac{\theta z}{12}\Big(1 - \frac{1}{36M^2}\Big) &\geq \frac{\Big(1 - \frac{1}{N-1}\Big)^2\Big(1 - \frac{2}{N}\Big)\Big(1 - \frac{1}{36M^2}\Big) \cdot N}{12 \cdot M^{(r-1)}} \\ &= \frac{\lambda N}{12 \cdot M^{(r-1)}} \ . \end{aligned} \tag{9}$$

From Equation (8) and Equation (9),

$$\Big(\theta e^{-z/12} + 1 - \theta\Big)^q \leq \exp\Big(\frac{-\lambda q N}{12 \cdot M^{r-1}}\Big) \ . \tag{10}$$

Analogously, we can show that

$$\Big(\theta e^{-z/9} + 1 - \theta\Big)^q \leq \exp\Big(\frac{-\lambda q N}{9 \cdot M^{(r-1)}}\Big) \ . \tag{11}$$

From Equation (7), Equation (10), and Equation (11), we obtain the claimed result.

Case 2: X and X' agree in the right segments. From Lemma 5.2, for any $s \in \mathbb{Z}_N$, the random variables $V_{1,s}, \ldots, V_{\ell,s}$ are independent and identically distributed, with $\Pr[V_{1,s} = 1] = p$. The chance the adversary RHR can correctly guess the right segment of X is at least the probability that $V_s \leq \ell(p + \Delta/2)$ for every $s \in \mathbb{Z}_N$. Let $z = \Delta^2/p$. Proceeding as in Case 1, we can show that for any $s \in \mathbb{Z}_N$,

$$\Pr[V_s \leq \ell(p + \Delta/2)] \leq e^{-z\ell/9} \ .$$

Let U be the random variable for the number of pairs (C,C') among q given pairs of ciphertexts such that C and C' agree on their left segments. From Lemma 5.2, U is a Binomial random variable $B(q, \theta)$, with

$$\theta \geq \frac{N-1}{MN-1} - \frac{1}{M \cdot (MN)^{(r-2)/2}} \geq \frac{1 - 1/(N-1)}{M};$$

where the last inequality exploits the hypothesis that $r \geq 6$. Hence the adversary can correctly guess the right segment of X with probability at least

$$\begin{aligned} 1 - \sum_{\ell=0}^{q} \Pr[U = \ell] \cdot \sum_{s \in \mathbb{Z}_N} \Pr[V_s \geq \ell(p + \Delta/2)]) &\geq 1 - N \cdot \sum_{\ell=0}^{q} \binom{q}{\ell} \theta^\ell (1-\theta)^{q-\ell} t e^{-z\ell/9} \\ &= 1 - N \cdot \Big(\theta e^{-z/9} + 1 - \theta\Big)^q \ . \end{aligned}$$

From Equation (11),

$$\Big(\theta e^{-z/9} + 1 - \theta\Big)^q \leq \exp\Big(\frac{-\lambda q N}{9 \cdot M^{r-1}}\Big) \ .$$

Hence the adversary can correctly guess the right segment of X with probability at least

$$1 - N \cdot \exp\Big(\frac{-\lambda q N}{9 \cdot M^{r-1}}\Big) \geq 1 - N \cdot \exp\Big(\frac{-\lambda q N}{9 \cdot M^{r-1}}\Big) - \exp\Big(\frac{-\lambda q N}{12 \cdot M^{r-1}}\Big) \ .$$

Proof of Lemma 6.2. Let $\mathrm{Match}_t(Z)$ be the event $(R_t \boxminus R'_t = Z) \wedge (L_t = L'_t)$. Recall that we have to study

$$\pi(Z) := \Pr[R_t \boxminus R'_t = Z \mid L_t = L'_t] = \frac{\Pr[\mathrm{Match}_t(Z)]}{\Pr[L_t = L'_t]} \ .$$

The denominator of the right-hand side has both upper and lower bounds by Lemma 5.2, and it doesn't depend on the value of Z. To bound the numerator, we'll need Lemma 6.3 below, which is the counterpart of Lemma 5.5 for LHR attack. A proof sketch of Lemma 6.3 is given further below. The full proof is in Appendix B.2.

Lemma 6.3 *Let* $\mathsf{F} = \mathbf{Feistel}[r, M, N, \boxplus]$. *Fix distinct* $X, X' \in \mathbb{Z}_M \times \mathbb{Z}_N$, *and* $T \in \mathsf{F.Twk}$, *an even integer* $t \in \{4, 5, \ldots, r\}$. *Pick* $K \leftarrow\!\!\$\ \mathsf{F.Keys}$. *Let* L_t *and* R_t *denote the left and right segment of the round-t output of* X *under* $\mathsf{F.E}(K, T, \cdot)$. *Define* L'_t *and* R'_t *for* X' *likewise. For each* $Z \in \mathbb{Z}_M \backslash \{0\}$, *let* $\mathrm{Match}_t(Z)$ *be the event* $(R_t \boxminus R'_t = Z) \wedge (L_t = L'_t)$. *Then, for any* $Z, Z' \in \mathbb{Z}_M \backslash \{0, R_0 \boxminus R'_0\}$,

$$\Pr[\mathrm{Match}_t(Z)] = \Pr[\mathrm{Match}_t(Z')] \ . \tag{12}$$

Moreover, if $R_0 \neq R'_0$ then for any $Z \in \mathbb{Z}_M \backslash \{0, R_0 \boxminus R'_0\}$,

$$\Pr[\mathrm{Match}_t(R_0 \boxminus R'_0)] \geq \Pr[\mathrm{Match}_t(Z)] + \frac{1}{M^{t/2}} \ . \tag{13}$$

Back to the proof of Lemma 6.2, first consider part (a). We recall that in this part, X and X' have the same right segment, meaning that $R_0 \boxminus R'_0 = 0$. From Lemma 5.2, $\Pr[L_t = L'_t] > 0$, and thus the conditional probability in the claimed result is well-defined. Lemma 6.3 claims that in this case, $\Pr[\mathrm{Match}_t(Z)]$ doesn't depend on the value of Z, for all $Z \in \mathbb{Z}_M \backslash \{0\}$, and thus $\pi(Z) = \frac{\Pr[\mathrm{Match}_t(Z)]}{\Pr[L_t = L'_t]}$ also doesn't depend on the value of Z. Moreover, if $L_t = L'_t$ then $R_t \neq R'_t$, because X and X' are distinct, and thus for any $Z \in \mathbb{Z}_M \backslash \{0\}$,

$$\pi(Z) = \frac{1}{N-1} \sum_{Z' \in \mathbb{Z}_N \backslash \{0\}} \pi(Z') = \frac{1}{N-1} \ .$$

We now prove the claims in part (b). Recall that in this part, the messages X and X' have different right segments. From Lemma 5.2, $\Pr[L_t = L'_t] > 0$, and thus the conditional probabilities in the our claims are well-defined. Again, from Lemma 6.3, $\Pr[\mathrm{Match}_t(Z)]$ doesn't depend on the value of Z, for all $Z \in \mathbb{Z}_M \backslash \{0, R_0 \boxminus R'_0\}$, and thus $\pi(Z) = \frac{\Pr[\mathrm{Match}_t(Z)]}{\Pr[L_t = L'_t]}$ also doesn't depend on the value of Z. Hence for any $Z \in \mathbb{Z}_M \backslash \{0, R_0 \boxminus R'_0\}$,

$$\pi(Z) = \tfrac{1}{N-2} \sum_{Z' \notin \{0, R_0 \boxminus R'_0\}} \pi(Z') = \tfrac{1}{N-2} \Big(1 - \pi(R_0 \boxminus R'_0)\Big) \ .$$

Therefore, if we can prove the claimed lower bound of $\pi(R_0 \boxminus R'_0)$ then the upper bound of $\pi(Z)$ will automatically follow. Fix $Z \in \mathbb{Z}_N \backslash \{0, R_0 \boxminus R'_0\}$. From Lemma 6.3,

$$\Pr[\mathrm{Match}_t(R_0 \boxminus R'_0)] \geq \Pr[\mathrm{Match}_t(Z)] + \frac{1}{M^{t/2}} \ .$$

Moreover, from Lemma 5.2,

$$\Pr[L_t = L'_t] \leq \frac{N-1}{MN-1} + \frac{1}{(MN)^{t/2}} \leq \frac{1}{M} \ .$$

Hence

$$\pi(R_0 \boxminus R'_0) - \pi(Z) = \frac{\Pr[\mathrm{Match}_t(R_0 \boxminus R'_0)] - \Pr[\mathrm{Match}_t(Z)]}{\Pr[L_t = L'_t]} \geq \frac{1}{M^{t/2} \cdot \Pr[L_t = L'_t]} \geq \frac{1}{M^{(t-2)/2}} \ .$$

Consequently, $1 - \pi(R_0 \boxminus R'_0)$ is upper bounded by

$$\sum_{Z \notin \{0, R_0 \boxminus R'_0\}} \pi(Z) \leq (N-2) \Big(\pi(R_0 \boxminus R'_0) - \frac{1}{M^{(t-2)/2}}\Big) \ ,$$

which can be resolved to

$$\pi(R_0 \boxminus R'_0) \geq \frac{1}{N-1} + \frac{1 - 1/(N-1)}{M^{(t-2)/2}} \ .$$

Proof sketch for Lemma 6.3. We'll show the following recursion: for every $V \in \mathbb{Z}_N \backslash \{0\}$ and for even $t \geq 4$,

$$\Pr[\mathrm{Match}_t(V)] = \tfrac{\Pr[\mathrm{Match}_{t-2}(V)]}{M} + \tfrac{\Pr[L_{t-2} \neq L'_{t-2}]}{MN} \ .$$

To see why this is true, note that $\text{Match}_t(V)$ can be reduced to the conjunction of (i) $L_{t-1} = L'_{t-1}$, and (ii) $R_{t-2} \boxminus R'_{t-2} = V$. If (ii) happens then $R_{t-2} \neq R'_{t-2}$, and consequently L_{t-1} and L'_{t-1} are uniform and independent. Hence $\Pr[\text{Match}_t(V)] = \frac{\Pr[R_{t-2} \boxminus R'_{t-2} = V]}{M}$. Next, $R_{t-2} \boxminus R'_{t-2} = V$ is the union of the disjoint events $\text{Match}_{t-2}(V)$ and (iii) $(R_{t-2} \boxminus R'_{t-2} = V) \wedge (L_{t-2} \neq L'_{t-2})$. Since $L_{t-3} = L_{t-2}$ and $L'_{t-3} = L'_{t-2}$, if $L_{t-2} \neq L'_{t-2}$ then $L_{t-3} \neq L'_{t-3}$, and consequently R_{t-2} and R'_{t-2} are independent and uniform. Hence (iii) happens with probability $\frac{\Pr[(L_{t-2} \neq L'_{t-2})]}{N}$.

To prove the claims in Lemma 6.3, we'll use an induction proof on t. For the base case $t = 2$, one can show that $\Pr[\text{Match}_2(Z)] = 0$ for $Z \in \mathbb{Z}_N \backslash \{0, R_0 \boxminus R'_0\}$, and $\Pr[\text{Match}_2(R_0 \boxminus R'_0)]$ is $\frac{1}{M}$ if $R_0 \neq R'_0$, and is 0 otherwise. The inductive step is as follows. To justify Equation (12), use the recursion above to compute $\Pr[\text{Match}_t(Z)]$ from $\Pr[\text{Match}_{t-2}(Z)]$, and $\Pr[\text{Match}_t(Z')]$ from $\Pr[\text{Match}_{t-2}(Z')]$, and then apply the induction hypothesis. For Equation (13), use the recursion above to compute $\Pr[\text{Match}_t(R_0 \boxminus R'_0)]$ from $\Pr[\text{Match}_{t-2}(R_0 \boxminus R'_0)]$, and $\Pr[\text{Match}_t(Z)]$ from $\Pr[\text{Match}_{t-2}(Z)]$, and then apply the induction hypothesis.

7 The Full-Message Recovery attack

Recall that the LHR attack recovers the left segment of the target message X, while the RHR attack recovers the right segment of X. By combining them, one can fully recover X as follows. We require ciphertexts of three messages X, X', X^* for q tweaks, with sufficiently large q, to recover X. The message X' is fully known but has no relation with the target X; this is already enough to recover the right segment of X, according to the RHR attack. The message X^* is required to have the same right segment as the target, but it's only partially known: only the left segment of X^* is included in the auxiliary information. For example, X^* is the "default" version of X, in which the left segment is 0. Although X^* is only partially known, as mentioned above, we already recovered the right segment of X (and also X^*). Then the LHR attack gives us the left segment of X.

THE ATTACK. Fix integer $q \geq 1$. Let $\mathsf{DC3}_q$ be the class of all algorithms D that output $(L', R') \leftarrow X' \in \mathbb{Z}_M \times \mathbb{Z}_N, L^* \in \mathbb{Z}_M \backslash \{L'\}$ and distinct tweaks $T_1, \ldots, T_q \in \mathsf{F.Twk}$. Let $\mathsf{SC3}_q = \{\mathsf{XS}[\mathsf{D}] \mid \mathsf{D} \in \mathsf{DC3}_q\}$, where each sampler $\mathsf{XS}[\mathsf{D}]$ in $\mathsf{SC3}_q$ behaves as follows

Sampler $\mathsf{XS}[\mathsf{D}]$
$(X', L^*, T_1, \ldots, T_q) \leftarrow\!\!\$\ \mathsf{D}\ ;\ (L', R') \leftarrow X'$
$L \leftarrow\!\!\$\ \mathbb{Z}_M \backslash \{L', L^*\}\ ;\ R \leftarrow\!\!\$\ \mathbb{Z}_N\ ;\ X \leftarrow (L, R)\ ;\ X^* \leftarrow (L^*, R)\ ;\ a \leftarrow (X', L^*)$
Return $((T_1, X^*), (T_1, X'), (T_1, X), \ldots, (T_q, X^*), (T_q, X'), (T_q, X), X, a)$

The sampler above picks the left segment L of the target X uniformly random, with the condition that L must be different from both L^* and the left segment L' of the known message X' produced by D. It then picks the right segment R of X uniformly, and let $X^* = (L^*, R)$ be another partially known message. The auxiliary information contains X' and the left half L^* of X^*. The number of examples is $Q = 3q$; the number of tweaks is $q_t = q$; the number of target tweaks is $q^* = q$; and the number of examples per tweak is $q_e = 3$. Since X, X', and X^* are distinct, each sampler in $\mathsf{SC3}_q$ satisfies the distinctness condition. The Full-Message Recovery attack FMR against $\mathsf{SC3}_q$ is shown in Fig. 12. Since q_e is small, we would expect and desire that adversaries have low mr-advantage, even if Q is big. Indeed, an ideal FPE scheme has this property. Our FMR attack shows that Feistel-based FPE fails to have this property. Theorem 7.1 below gives a lower bound on the mr advantage of FMR; this bound is illustrated in Fig. 13.

Adversary $\mathsf{FMR}((T_1, C_1^*), (T_1, C_1'), (T_1, C_1), \ldots, (T_q, C_q^*), (T_q, C_q'), (T_q, C_q), a)$

$(X', L^*) \leftarrow a$; $(L', R') \leftarrow X'$

$a_1 \leftarrow (L^*, R')$; $X^* \leftarrow \mathsf{RHR}((T_1, C_1'), (T_1, C_1), \ldots, (T_q, C_q'), (T_q, C_q), a_1)$

$a_2 \leftarrow X^*$; $X \leftarrow \mathsf{LHR}((T_1, C_1^*), (T_1, C_1), \ldots, (T_q, C_q^*), (T_q, C_q), a_2)$

Return X

Figure 12: **The Full-Message Recovery attack.**

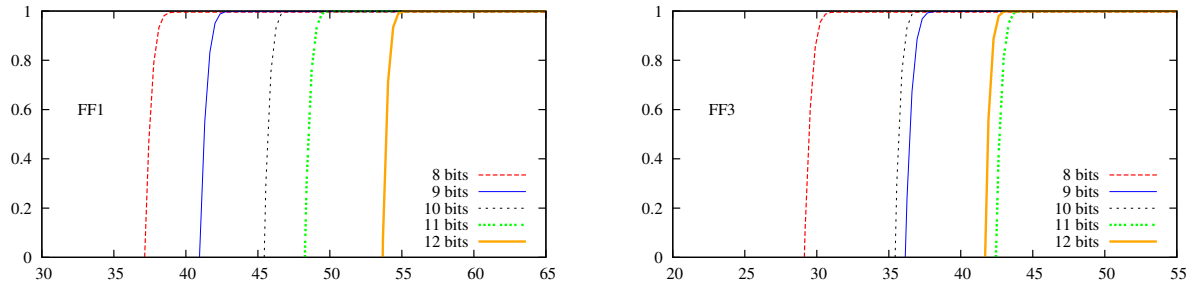

Figure 13: **The mr advantage of the Full-Message Recovery attack for binary strings of 8–12 bits.** The x-axis shows the log, base 2, of the number q of ciphertext triples, and the y-axis shows $\mathbf{Adv}^{\mathsf{mr}}_{\mathbf{Feistel}[r,M,N,\boxplus],\mathsf{XS}}(\mathsf{FMR})$, for $\mathsf{XS} \in \mathsf{SC3}_q$. On the left, we use the parameters of the FF1 standard. On the right, we use parameters of FF3.

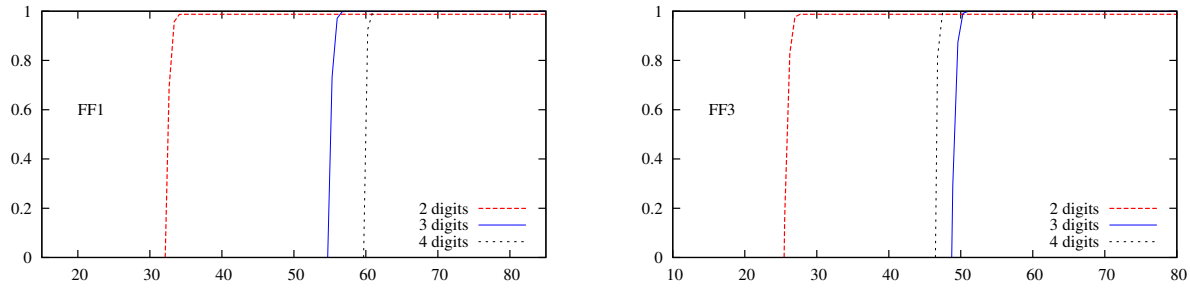

Figure 14: **The mr advantage of the Full-Message Recovery attack fordecimal strings of 2–4 digits.** The x-axis shows the log, base 2, of the number q of ciphertext triples, and the y-axis shows $\mathbf{Adv}^{\mathsf{mr}}_{\mathbf{Feistel}[r,M,N,\boxplus],\mathsf{XS}}(\mathsf{FMR})$, for $\mathsf{XS} \in \mathsf{SC3}_q$. On the left, we use the parameters of the FF1 standard, meaning that $r = 10$, and for ℓ-bit strings, $M = 2^{\lfloor \ell/2 \rfloor}$ and $N = 2^{\lceil \ell/2 \rceil}$. On the right, we use parameters of FF3, meaning that $r = 8$, and for ℓ-bit strings, $M = 2^{\lceil \ell/2 \rceil}$ and $N = 2^{\lfloor \ell/2 \rfloor}$.

Theorem 7.1 *Let $N, M \geq 3$ and $q \geq 1$ be integers, and let $r \geq 6$ be an even integer such that $M^{(r-2)/2} \geq 2N$ and $N^{(r-2)/2} \geq 2M$. Let $\lambda_1 = \left(1 - \frac{1}{N-1}\right)^2 \left(1 - \frac{2}{N}\right)\left(1 - \frac{1}{36M^2}\right)$ and $\lambda_2 = \left(1 - \frac{1}{M-1}\right)^2 \left(1 - \frac{1}{MN}\right)$. Let $\mathsf{F} = \mathbf{Feistel}[r, M, N, \boxplus]$. Then for any sampler XS in the class $\mathsf{SC3}_q$,*

$$\mathbf{Adv}^{\mathsf{mr}}_{\mathsf{F},\mathsf{XS}}(\mathsf{FMR}) \quad \geq \quad 1 - N \cdot \exp\left(\frac{-\lambda_1 q N}{9 \cdot M^{r-1}}\right) - M \cdot \exp\left(\frac{-\lambda_2 M p}{9 \cdot N^{r-2}}\right)$$

$$- \exp\Big(\frac{-\lambda_2 Mp}{12 \cdot N^{r-2}}\Big) - \exp\Big(\frac{-\lambda_1 qN}{12 \cdot M^{r-1}}\Big) - \frac{1}{N(M-2)} \ .$$

Proof: First we'll show that $\mathbf{Adv}^{\mathsf{mg}}_{\mathsf{XS}} \leq \frac{1}{N(M-2)}$. Consider an arbitrary simulator $\mathcal{S}$. The simulator is given $(L', R') \in \mathbb{Z}_M \times \mathbb{Z}_N$ and $L^* \in \mathbb{Z}_M$, and has to guess (L, R), where $R \leftarrow\!\!\$\ \mathbb{Z}_N$ and $L \leftarrow\!\!\$\ \mathbb{Z}_M \backslash \{L^*, L'\}$. Hence $\Pr[\mathbf{G}^{\mathsf{mg}}_{\mathsf{XS}}(\mathcal{S})] \leq \frac{1}{N(M-2)}$. Since this bound holds for any simulator,

$$\mathbf{Adv}^{\mathsf{mg}}_{\mathsf{XS}} = \max_{\mathcal{S}} \Pr[\mathbf{G}^{\mathsf{mg}}_{\mathsf{XS}}(\mathcal{S})] \leq \frac{1}{N(M-2)} \ .$$

We now show that $\Pr[\mathbf{G}^{\mathsf{mr}}_{\mathsf{F},\mathsf{XS}}(\mathsf{FMR})] \geq 1 - N \cdot \exp\Big(\frac{-\lambda_1 qN}{9 \cdot M^{r-1}}\Big) - M \cdot \exp\Big(\frac{-\lambda_2 Mp}{9 \cdot N^{r-2}}\Big) - \exp\Big(\frac{-\lambda_2 Mp}{12 \cdot N^{r-2}}\Big) - \exp\Big(\frac{-\lambda_1 qN}{12 \cdot M^{r-1}}\Big)$. The adversary FMR first calls RHR on the ciphertexts of X and X', and the auxiliary information gives X and L^*. Although L^* is *not* the left segment of the target message X, from the proof of Theorem 6.1, adversary RHR still can recover the right segment R of X (and thus what it outputs is $X^* = (L^*, R)$) with probability at least $1 - N \cdot \exp\Big(\frac{-\lambda_1 qN}{9 \cdot M^{r-1}}\Big) - \exp\Big(\frac{-\lambda_1 qN}{12 \cdot M^{r-1}}\Big)$. Next, if what RHR outputs is X^* then from the proof of Theorem 5.1, adversary LHR can correctly output X with probability at least $1 - M \cdot \exp\Big(\frac{-\lambda_2 Mp}{9 \cdot N^{r-2}}\Big) - \exp\Big(\frac{-\lambda_2 Mp}{12 \cdot N^{r-2}}\Big)$. By union bound, FMR can recover X with probability at least $1 - N \cdot \exp\Big(\frac{-\lambda_1 qN}{9 \cdot M^{r-1}}\Big) - M \cdot \exp\Big(\frac{-\lambda_2 Mp}{9 \cdot N^{r-2}}\Big) - \exp\Big(\frac{-\lambda_2 Mp}{12 \cdot N^{r-2}}\Big) - \exp\Big(\frac{-\lambda_1 qN}{12 \cdot M^{r-1}}\Big)$. ∎

Acknowledgments

We thank the CCS reviewers for their insightful comments.

References

[1] M. Bellare, T. Ristenpart, P. Rogaway, and T. Stegers. Format-preserving encryption. In M. J. Jacobson Jr., V. Rijmen, and R. Safavi-Naini, editors, *SAC 2009*, volume 5867 of *LNCS*, pages 295–312. Springer, Heidelberg, Aug. 2009. 3, 4, 5, 6, 10, 14, 29

[2] M. Bellare and P. Rogaway. The security of triple encryption and a framework for code-based game-playing proofs. In S. Vaudenay, editor, *EUROCRYPT 2006*, volume 4004 of *LNCS*, pages 409–426. Springer, Heidelberg, May / June 2006. 7

[3] M. Bellare, P. Rogaway, and T. Spies. Addendum to "the FFX mode of operation for Format-Preserving Encryption". Draft 1.0. Submission to NIST, Sept. 2010. `http://csrc.nist.gov/groups/ST/toolkit/BCM/documents/proposedmodes/ffx/ffx-spec2.pdf`. 3, 5

[4] M. Bellare, P. Rogaway, and T. Spies. The FFX mode of operation for format-preserving encryption. Draft 1.1. Submission to NIST, Feb. 2010. `http://csrc.nist.gov/groups/ST/toolkit/BCM/documents/proposedmodes/ffx/ffx-spec.pdf`. 3, 5

[5] J. Black and P. Rogaway. Ciphers with arbitrary finite domains. In B. Preneel, editor, *CT-RSA 2002*, volume 2271 of *LNCS*, pages 114–130. Springer, Heidelberg, Feb. 2002. 3

[6] E. Brier, T. Peyrin, and J. Stern. BPS: a format-preserving encryption proposal. Submission to NIST, 2010. `http://csrc.nist.gov/groups/ST/toolkit/BCM/documents/proposedmodes/bps/bps-spec.pdf`. 3, 5

Game $\mathbf{G}^{\mathsf{mr1}}_{\mathsf{F}}(\mathcal{A})$	Game $\mathbf{G}^{\mathsf{mg1}}_{\mathcal{A}}(\mathcal{S})$
$K \leftarrow\!\!\$\ \mathsf{F.Keys}$; $(T, X) \leftarrow\!\!\$\ \mathcal{A}$; $n \leftarrow 0$	$(T, X) \leftarrow\!\!\$\ \mathcal{A}$; $n \leftarrow 0$
$Y \leftarrow \mathsf{F.E}(K, T, X)$; $X^* \leftarrow\!\!\$\ \mathcal{A}^{\text{Enc}}(T, Y)$	$X^* \leftarrow\!\!\$\ \mathcal{S}^{\text{Eq}}(T)$
Return $(X^* = X)$	Return $(X^* = X)$
$\underline{\text{Enc}(T', X')}$	$\underline{\text{Eq}(X')}$
$n \leftarrow n + 1$; $(T_n, X_n) \leftarrow (T', X')$; $Y_n \leftarrow \mathsf{F.E}(K, T_n, X_n)$	$n \leftarrow n + 1$
Return Y_n	Return $(X' = X)$

Figure 15: **Games for the MR1 message-recovery security definition of FPE scheme F.**

[7] M. Dworkin. Recommendation for Block Cipher Modes of Operation: Methods for Format-Preserving Encryption. *NIST Special Publication 800-38G*, Mar. 2016. `http://dx.doi.org/10.6028/NIST.SP.800-38G`. 3, 4, 5, 7, 8

[8] M. Liskov, R. L. Rivest, and D. Wagner. Tweakable block ciphers. *Journal of Cryptology*, 24(3):588–613, July 2011. 3, 5, 10, 18

[9] J. Patarin. New results on pseudorandom permutation generators based on the DES scheme. In J. Feigenbaum, editor, *CRYPTO'91*, volume 576 of *LNCS*, pages 301–312. Springer, Heidelberg, Aug. 1992. 4, 6, 13, 14, 18

[10] J. Patarin. Generic attacks on Feistel schemes. In C. Boyd, editor, *ASIACRYPT 2001*, volume 2248 of *LNCS*, pages 222–238. Springer, Heidelberg, Dec. 2001. 6, 18

[11] J. Patarin. Security of random Feistel schemes with 5 or more rounds. In M. Franklin, editor, *CRYPTO 2004*, volume 3152 of *LNCS*, pages 106–122. Springer, Heidelberg, Aug. 2004. 4, 6, 18

A MR1 message recovery definition

Let F be an FPE scheme. We recall the BRRS [1] definition of message recovery security, adapted to our notation, calling it MR1. First consider game mr1 on the left of Fig. 15, associated to F and adversary $\mathcal{A}$. The latter is first executed to obtain a target tweak T and message X. The latter is encrypted under K relative to T to get ciphertext Y, and $\mathcal{A}$ is now run on inputs T, Y, winning if the guess X^* it returns equals X. (The two executions of $\mathcal{A}$ are entirely independent, sharing neither coins nor state. The first execution is a way for $\mathcal{A}$ to specify the distribution of T, X.) In its second execution, $\mathcal{A}$ can obtain input-output examples under K via its Enc oracle which takes a tweak T' and message X' to return the corresponding ciphertext, modeling a classical chosen-plaintext attack. Assume $\mathcal{A}$ makes q queries to this oracle. Then its advantage is

$$\mathbf{Adv}^{\mathsf{mr1}}_{\mathsf{F}}(\mathcal{A}) = \Pr[\mathbf{G}^{\mathsf{mr1}}_{\mathsf{F}}(\mathcal{A})] - \max_{\mathcal{S}} \Pr[\mathbf{G}^{\mathsf{mg1}}_{\mathcal{A}}(\mathcal{S})] \ ,$$

where the maximum is over all simulators $\mathcal{S}$ making q queries to their Eq oracle in the mg1 game on the right of Fig. 15. In this game, the target tweak T and message X are selected, just like in game mr1, by running $\mathcal{A}$, but no encryption is performed and no ciphertext is provided to $\mathcal{S}$. Instead the latter gets an equality oracle Eq to which it can submit any message X', learning in response whether or not X' equals X.

The intuition here is that $\mathcal{A}$ can test whether a candidate X' equals X by callling $\text{Enc}(T, X')$ and seeing whether the result equals Y. The simulator thus gets the same ability via its Eq oracle. Note that accordingly the number of Eq queries allowed to $\mathcal{S}$ is exactly the number q of Enc queries made by $\mathcal{A}$. We now discuss the definition and compare it with ours.

The classical choice of allowing the adversary a chosen-plaintext attack that mr1 incorporates results in a strong definition, a plus when one can prove schemes meet it, but a minus with regard to classifying attacks. Consider two attacks, each using q input-output examples, in one of which the adversary needs to know nearly nothing about the example plaintexts, and the other in which it needs to know them in their entirety, both having the same probability of recovering the target message. Both can be cast as mr1 adversaries, but since the examples are obtained via ENC, the first attack is forced to pick some messages in its queries, so that it knows them. The difference between the attacks, which is important in practice, does not surface in the formal claim that would be made about their mr1 advantage. However, in our framework, the difference surfaces, as the advantages would be relative to different classes of message samplers.

Let $N = |\mathsf{F.Dom}|$ and consider any $\mathcal{A}$ making $q = N$ queries to ENC. Then, under the mr1 definition, $\mathcal{A}$ will have zero (more precisely, non-positive) advantage, meaning $\mathbf{Adv}^{\mathsf{mr1}}_{\mathsf{F}}(\mathcal{A}) \leq 0$. This is because $\mathcal{S}$ is now allowed $q = N$ queries to EQ. It can query all points in $\mathsf{F.Dom}$, and will thereby certainly find X. So according to the mr1 definition, attacks with $q \geq N$ are not viewed as successful. But in practice, they can be damaging, depending on the example tweaks and messages involved. For example, suppose there were an attack that, given encryptions $Y_1, \ldots, Y_q$ of a single example message Z under $q \geq N$ different tweaks $T_1, \ldots, T_q$, recovers X. This is not something we would like, and there is no reason a well-designed FPE scheme should be subject to such an attack. But if all we know is that it meets mr1, such an attack may well exist. In our framework, however, such an attack would have a high mr advantage relative to a certain class of message samplerrs, meaning the framework correctly evaluates the attack as successful.

One could address the second point by altering MR1 so that the number of EQ queries of $\mathcal{S}$ is the number of $\text{ENC}(T, \cdot)$ queries of $\mathcal{A}$, meaning the number in which the tweak equals the target one. Still, the definition is conceptually different because it focuses on the number of queries or examples rather than what these are. (The simulator's EQ queries could be on messages entirely different from those in the adversary's ENC queries.) In our definition, the adversary and simulator are given exactly the same examples, reflecting a tighter connection.

B Deferred proofs

B.1 Proof of Lemma 5.2

We'll prove the claim in part (a); the proof for part (b) is similar. The idea of the proof is to give a recursion to compute $\Pr[L_t = L'_t]$ from $\Pr[L_{t-2} = L'_{t-2}]$, and then give an induction proof on t.

Let F_i be the round function of F at round i, and let $G_i(\cdot,\cdot)$ be $F_i(K,\cdot,\cdot)$. Let R_t be the the right segment of the round-t output of X under $\mathsf{F}(K,T,\cdot)$. Define R'_t for X' likewise. Note that $L_t = L_{t-1} = G_{t-1}(T, R_{t-2}) \boxplus L_{t-2}$ and $L'_t = L'_{t-1} = G_t(T, R'_{t-2}) \boxplus L'_{t-2}$.

We shall prove by induction on t. First consider the base case $t = 2$. Since $R_0 = R'_0$ and $L_0 \neq L'_0$,

$$L_2 = G_1(T, R_0) \boxplus L_0 \neq G_1(T, R'_0) \boxplus L'_0 = L'_2 \ .$$

Hence the claim holds for the base case. Suppose that the claims also hold for $t-2$. We shall prove that they hold for t as well. Since $\mathbf{Feistel}[t-2, M, N, \boxplus]$ is an FPE scheme and $X \neq X'$, we must have $(L_{t-2}, R_{t-2}) \neq (L'_{t-2}, R'_{t-2})$. On the one hand, if $R_{t-2} = R'_{t-2}$ then L_{t-2} and L'_{t-2} must be different, thus

$$L_t = G_{t-1}(T, R_{t-2}) \boxplus L_{t-2} \neq G_{t-1}(T, R'_{t-2}) \boxplus L'_{t-2} = L'_t \ .$$

In other words,

$$\Pr[L_t = L'_t \mid R_{t-2} = R'_{t-2}] = 0 \ . \tag{14}$$

On the other hand, if $R_{t-2} \neq R'_{t-2}$, since G_t is a truly random function from $\mathbb{Z}_N$ to $\mathbb{Z}_M$, the conditional probability that

$$G_{t-1}(T, R_{t-2}) \boxplus L_{t-2} = G_{t-1}(T, R'_{t-2}) \boxplus L'_{t-2}$$

is exactly $1/M$, and thus

$$\Pr[L_t = L'_t \mid R_{t-2} \neq R'_{t-2}] = \frac{1}{M} \ . \tag{15}$$

From Equation (14) and Equation (15),

$$\begin{aligned} \Pr[L_t = L'_t] &= \frac{1}{M}(1 - \Pr[R_{t-2} = R'_{t-2}]) \\ &= \frac{1}{M}(1 - \Pr[R_{t-1} = R'_{t-1}]) \ . \end{aligned}$$

By symmetry,

$$\Pr[R_{t-1} = R'_{t-1}] = \frac{1}{N}(1 - \Pr[L_{t-2} = L'_{t-2}]) \ .$$

Hence

$$\Pr[L_t = L'_t] = \frac{N-1}{MN} + \frac{1}{MN} \Pr[L_{t-2} = L'_{t-2}] \ . \tag{16}$$

From Equation (16), using the induction hypothesis yields the claimed bounds for t.

B.2 Proof of Lemma 6.3

We'll show that for any even $t \geq 2$ and any $V \in \mathbb{Z}_N \backslash \{0\}$,

$$\Pr[\mathrm{Match}_t(V)] = \frac{\Pr[R_{t-2} \boxminus R'_{t-2} = V]}{M}, \tag{17}$$

and if $t \geq 4$ then

$$\Pr[\mathrm{Match}_t(V)] = \frac{\Pr[\mathrm{Match}_{t-2}(V)]}{M} + \frac{\Pr[L_{t-2} \neq L'_{t-2}]}{MN} \tag{18}$$

The proofs of those claims are deferred further below; we now justify the claimed results of this lemma, namely Equation (12) and Equation (13). Fix Z and Z' in $\mathbb{Z}_N \backslash \{0, R_0 \boxminus R'_0\}$. We'll prove Equation (12) by induction on t. First consider the base case $t = 2$. From Equation (17),

$$\Pr[\mathrm{Match}_2(Z)] = \frac{1}{M} \cdot \Pr[R_0 \boxminus R'_0 = Z] = 0,$$

and likewise, $\mathrm{Match}_2(Z') = 0$. Hence Equation (12) holds for the base case. Now suppose that it holds for $t-2$. We'll prove that it also holds for t. From Equation (18),

$$\Pr[\mathrm{Match}_t(Z)] - \Pr[\mathrm{Match}_t(Z')] = \frac{1}{M}\Big(\Pr[\mathrm{Match}_{t-2}(Z)] - \Pr[\mathrm{Match}_{t-2}(Z')]\Big) \ .$$

From the induction hypothesis,

$$\Pr[\mathrm{Match}_{t-2}(Z)] = \Pr[\mathrm{Match}_{t-2}(Z')] \ .$$

Hence Equation (12) also holds for t.

We now justify Equation (13). Recall that in this case, $R_0 \neq R'_0$. Fix $Z \in \mathbb{Z}_N \backslash \{0, R_0 \boxminus R'_0\}$. We'll prove by induction on t. First consider the base case $t = 2$. From Equation (17),

$$\Pr[\mathrm{Match}_2(R_0 \boxminus R'_0)] = \frac{1}{M} \cdot \Pr[R_0 \boxminus R'_0 = R_0 \boxminus R'_0] = \frac{1}{M},$$

whereas

$$\Pr[\mathrm{Match}_2(Z)] = \frac{1}{M} \cdot \Pr[R_0 \boxminus R'_0 = Z] = 0 \ .$$

Hence Equation (13) holds for the base case $t = 2$. Now suppose that it holds for $t - 2$. From Equation (18),

$$\Pr[\mathrm{Match}_t(R_0 \boxminus R'_0)] - \Pr[\mathrm{Match}_t(Z)] = \frac{1}{M}\Big(\Pr[\mathrm{Match}_{t-2}(R_0 \boxminus R'_0)] - \Pr[\mathrm{Match}_{t-2}(Z)]\Big) \ .$$

Moreover, from the induction hypothesis,

$$\Pr[\mathrm{Match}_{t-2}(R_0 \boxminus R'_0)] - \Pr[\mathrm{Match}_{t-2}(Z)] \geq \frac{1}{M^{(t-2)/2}} \ .$$

Hence Equation (13) also holds for t.

What's left is to prove Equation (17) and Equation (18). Fix $V \in \mathbb{Z}_N \backslash \{0\}$. Let F_i be the round function of F at round i, let $G_i(\cdot, \cdot)$ be $F_i(K, \cdot, \cdot)$. Note that $R_t = G_t(T, L_{t-1}) \boxplus R_{t-1}$ and $R'_t = G_t(T, L'_{t-1}) \boxplus R'_{t-1}$. Since $L_{t-1} = L_t$ and $L'_{t-1} = L'_t$, if $L_t = L'_t$ then $R_t \boxminus R'_t = R_{t-1} \boxminus R'_{t-1}$, and thus

$$\Pr[\mathrm{Match}_t(V)] = \Pr[(R_{t-1} \boxminus R'_{t-1} = V) \wedge (L_t = L'_t)] \ .$$

Since $R_{t-1} = R_{t-2}$ and $R'_{t-1} = R'_{t-2}$,

$$\Pr[\mathrm{Match}_t(V)] = \Pr[(R_{t-2} \boxminus R'_{t-2} = V) \wedge (L_t = L'_t)] \ .$$

If $R_{t-2} \boxminus R'_{t-2} = V$ then $R_{t-2} \neq R'_{t-2}$, and thus

$$\Pr[L_t = L'_t \mid R_{t-1} \boxminus R'_{t-1} = V] = \frac{1}{M},$$

because (i) $L_t = L_{t-1} = G_{t-1}(T, R_{t-2}) \boxplus L_{t-2}$, (ii) $L'_t = L'_{t-1} = G_{t-1}(T, R'_{t-2}) \boxplus L'_{t-2}$, and (iii) G_{t-1} is independent of R_{t-2} and R'_{t-2}. Hence

$$\Pr[\mathrm{Match}_t(V)] \ = \ \frac{1}{M} \cdot \Pr[R_{t-2} \boxminus R'_{t-2} = V],$$

justifying Equation (17). To justify Equation (18), from Equation (17), it suffices to show that

$$\Pr[R_{t-2} \boxminus R'_{t-2} = V] = \Pr[\mathrm{Match}_{t-2}(V)] + \frac{\Pr[L_{t-2} \neq L'_{t-2}]}{N} \ .$$

Note that the event $R_{t-2} \boxminus R'_{t-2} = V$ is the union of the exclusive events $\mathrm{Match}_{t-2}(V)$ and $(R_{t-2} \boxminus R'_{t-2} = V) \wedge (L_{t-2} \neq L'_{t-2})$. If $L_{t-2} \neq L'_{t-2}$ then $L_{t-3} \neq L'_{t-3}$, because $L_{t-3} = L_{t-2}$ and $L'_{t-2} = L'_{t-3}$, and consequently, R_{t-2} and R'_{t-2} are uniformly random and independent. Hence

$$\Pr[(R_{t-2} \boxminus R'_{t-2} = V) \wedge (L_{t-2} \neq L'_{t-2})] = \frac{1}{N} \cdot \Pr[L_{t-2} \neq L'_{t-2}],$$

and thus

$$\Pr[R_{t-2} \boxminus R'_{t-2} = V] = \Pr[\mathrm{Match}_{t-2}(V)] + \frac{1}{N} \cdot \Pr[L_{t-2} \neq L'_{t-2}] \ .$$