Attacks Only Get Better:
How to Break FF3 on Large Domains

Viet Tung Hoang[1], David Miller[1], and Ni Trieu[2]

[1] Dept. of Computer Science, Florida State University, USA.
[2] Dept. of Computer Science, Oregon State University, USA.

Abstract. We improve the attack of Durak and Vaudenay (CRYPTO'17) on NIST Format-Preserving Encryption standard FF3, reducing the running time from $O(N^5)$ to $O(N^{17/6})$ for domain $\mathbb{Z}_N \times \mathbb{Z}_N$. Concretely, DV's attack needs about 2^{50} operations to recover encrypted 6-digit PINs, whereas ours only spends about 2^{30} operations. In realizing this goal, we provide a pedagogical example of how to use distinguishing attacks to speed up slide attacks. In addition, we improve the running time of DV's known-plaintext attack on 4-round Feistel of domain $\mathbb{Z}_N \times \mathbb{Z}_N$ from $O(N^3)$ time to just $O(N^{5/3})$ time. We also generalize our attacks to a general domain $\mathbb{Z}_M \times \mathbb{Z}_N$, allowing one to recover encrypted SSNs using about 2^{50} operations. Finally, we provide some proof-of-concept implementations to empirically validate our results.

Keywords: Format-preserving encryption, attacks

1 Introduction

Format-Preserving Encryption (FPE) [6, 12] is a form of deterministic symmetric encryption mechanism that preserves the *format* of plaintexts. For example, encrypting a 16-digit credit-card number under FPE would result in a 16-digit number, and encrypting a valid SSN would produce a ciphertext of nine decimal digits. FPE is widely used in practice by several companies, such as HPE Voltage, Verifone, Protegrity, Ingenico, to encrypt credit-card numbers and protect legacy databases. Recent research [4, 15, 20] however show that existing FPE standards FF1 and FF3 (NIST SP 800-38G, ANSI ASC X9.124) are somewhat vulnerable in small domains. The most damaging attack, due to Durak and Vaudenay (DV) [15], can recover the entire codebook of FF3 using $O(N^5)$ expected time, for domain $\mathbb{Z}_N \times \mathbb{Z}_N$.

Still, the attacks above are feasible only if the domain size is small; their cost becomes prohibitive for moderate and large domains. For example, for domain $\mathbb{Z}_{10}^6$ (namely encrypting 6-digit PINs), DV's attack would use about 2^{50} operations. In this paper, we improve DV's attack to break FF3 on large domains. Our attack can reduce the cost of breaking FF3 on domain $\mathbb{Z}_N \times \mathbb{Z}_N$ to $O(N^{17/6})$

N	**Our queries**	**DV's queries**	**Our rate**	**DV's rate**	**Our time**	**DV's time**
128	16,384	17,388	39%	56.85%	2^{20}	2^{35}
256	52,012	55,176	50%	55.9%	2^{23}	2^{40}
512	165,140	175,164	33%	77.4%	2^{26}	2^{45}

Table 1. Our attack versus DV's. The first column indicates the values of N in the domain $\mathbb{Z}_N \times \mathbb{Z}_N$. The second column and third column show the number of queries in our attack and that of DV respectively; in both attacks, the queries are made over two tweaks. The fourth and fifth columns show our recovery rate and that of DV respectively, and the fifth and sixth columns show our time and DV's time respectively.

expected time, meaning that it will need about 2^{30} operations to break FF3 of the domain $\mathbb{Z}_{10}^6$ above. Achieving this efficiency involves an elegant paradigm of combining distinguishing attacks with slide attacks [10, 11], and improved cryptanalyses of 4-round Feistel. We give rigorous analyses to justify the advantage of our attack, and provide proof-of-concept implementations in Section 5 that empirically confirm our analyses.

We note that our attack essentially performs the same queries as DV's, and thus the two attacks have the same scenario and asymptotic data/space complexity $\Theta(N^{11/6})$ for domain $\mathbb{Z}_N \times \mathbb{Z}_N$. However, DV use more aggressive choices of the parameters, and thus our attack is concretely better in both data and space complexity, albeit at the cost of lower recovery rate. A concrete comparison of the two attacks are given in Table 1. Still, one can improve the recovery rate by relaunching our attack with different tweaks. For example, for domain $\mathbb{Z}_{128} \times \mathbb{Z}_{128}$, if one relaunches our attack another time, the recovery rate would become $1 - (1 - 0.39)^2 \approx 62\%$. See Section 3.3 for further details.

Existing cryptanalysis. Let us begin by reviewing prior attacks on the standards FF1 and FF3. Bellare, Hoang, and Tessaro (BHT) [4] give the first attack on these schemes, showing that one can fully recover a target message using $O(N^6 \log(N))$ pairs of plaintext/ciphertext, on domain $\mathbb{Z}_N \times \mathbb{Z}_N$. Their attack however requires that a designated, partially known message must have the same right half as the target, but it is unclear how one could mount such a correlation in practice. Hoang, Tessaro, and Trieu (HTT) [20] subsequently improve BHT's attack, requiring no correlation between the known messages and the target. Even better, they can reuse the known plaintext/ciphertext pairs to attack multiple targets, thus reduce the *amortized* cost to $O(N^5 \log^2(N))$ pairs per target. Both attacks above apply to a *generic* Feistel-based FPE, meaning that they break both FF1 and FF3, and the only way to thwart them is to increase the round count of the underlying Feistel networks.

In a different direction, Durak and Vaudenay (DV) [15] give a dedicated attack on FF3, exploiting a bug in its design of round functions. They show that on domain $\mathbb{Z}_N \times \mathbb{Z}_N$, one can recover the entire codebook of FF3 using $O(N^{11/6})$ pairs of chosen plaintext/ciphertext, within $O(N^5)$ expected running time. We

stress that DV's attack does not apply to FF1, and it can be fixed without hurting performance by restricting the tweak space, as DV already suggested.

In response to DV's attack, NIST has temporarily suspended the use of FF3, whereas a draft update of the ANSI ASC X9.124 standard additionally recommends using double encryption on small domains to cope with the other attacks.

A BIRD'S-EYE VIEW ON DV'S ATTACK. We now briefly sketch a blueprint of DV's attack. Recall that in the balanced setting, the encryption scheme of FF3 is simply a tweakable blockcipher $\mathsf{F.E} : \mathsf{F.Keys} \times \mathsf{F.Twk} \times (\mathbb{Z}_N \times \mathbb{Z}_N) \to (\mathbb{Z}_N \times \mathbb{Z}_N)$ that is based on an 8-round balanced Feistel network. Due to a bug in the round functions of FF3, one can find two tweaks T and T^* such that $\mathsf{F.E}(K, T, \cdot)$ is the cascade $g(f(\cdot))$ of two 4-round Feistel networks f and g, whereas $\mathsf{F.E}(K, T^*, \cdot) = f(g(\cdot))$. Then, by mounting a slide attack using $O(N^{11/6})$ encryption queries, we obtain $O(N^2)$ instances, each of $O(N^{5/3})$ pairs of plaintext/ciphertext for f and also $O(N^{5/3})$ pairs for g. However, often just one of those instances provides the correct ciphertexts under f or g; in the remaining instances, the ciphertexts are random strings, independent of the plaintexts. DV resolve this by developing a codebook-recovery attack on 4-round Feistel networks using $O(N^3)$ expected running time. They then try this attack on every instance, using totally $O(N^5)$ expected time.

CONTRIBUTION: ELIMINATING FALSE INSTANCES. To improve the running time of DV's attack, we observe that it is an overkill to use an expensive codebook-recovery attack on false instances. A better solution is to find a cheap test to tell whether an instance is true or false, and then use the codebook-recovery attack on the true instances. A natural choice for such a test is a distinguishing attack on 4-round Feistel. However, the requirement here is a lot more stringent. To eliminate most of the random instances, our distinguishing attack should output 1 with probability about $1/N$ if it is given a false instance. To ensure that we will *not* incorrectly eliminate all true instances, the distinguishing attack should output 1 with high probability, say $1/2$, if it is given a true instance.

Our starting point is Patarin's distinguishing attack on 4-round Feistel [25],[3] which uses $O(\sqrt{N})$ pairs of plaintext/ciphertext. However, using this attack for our purpose runs into two obstacles. First, Patarin's asymptotic analysis is insufficient to pinpoint the hidden constant in the Big-Oh. Next, Patarin's attack fails to meet the requirement above, as given a false instance, the attack outputs 1 with constant probability.

Given the issues above, we instead design a new distinguishing attack, Left-Half Differential (LHD), such that (1) in the ideal world, it returns 1 with probability at most $\frac{1}{\sqrt{N}}$, and (2) in the real world, it returns 1 with probability at least $1 - \frac{1}{8\sqrt{N}} - \frac{10}{N} - \frac{1}{N^{3/4}}$. The LHD attack uses $O(N^{5/6})$ pairs of plaintext/ciphertext, and runs in $O(N^{5/6})$ time. Our analyses are generalized enough to include Patarin's attack as a special case. As a result, we can show that for $N \geq 2^{16}$, if one uses

[3] While Patarin's attack is given for classic Feistel (meaning that $N = 2^n$, and the underlying operator is xor), generalizing it to cover FF3 setting is straightforward.

Type	Power	Source	Data	Time
Chosen plaintext	Distinguishing	[24, 1] Here	$O(\sqrt{N})$	$O(\sqrt{N})$
Known-plaintext	Full recovery	[15]	$O(N^{5/3})$	$O(N^3)$
Known-plaintext	Full recovery	Here	$O(N^{5/3})$	$O(N^{5/3})$
Chosen plaintext & ciphertext	Full recovery	[9]	$O(N^{3/2})$	$O(N^{3/2})$

Table 2. A list of attacks on generic 4-round Feistel of domain $\mathbb{Z}_N \times \mathbb{Z}_N$. While the distinguishing attack was discovered by Patarin [24] and independently by Aiello and Venkatesan [1], the analyses in those papers are asymptotic. Our paper gives the first concrete treatment for this attack.

$\lceil 7 \cdot \sqrt{N} \rceil$ pairs of plaintext/ciphertext then Patarin's attack achieves advantage at least $1/2$.

In our test, we run LHD *twice*, first on the plaintext/ciphertext pairs of f, and then on those of g. Thus given a false instance, the chance that we fail to eliminate it is at most $\frac{1}{N}$, whereas given a true instance, the chance that we accept it is at least $\left(1 - \frac{1}{8\sqrt{N}} - \frac{10}{N} - \frac{1}{N^{3/4}}\right)^2$. Even better, our experiments indicate that in practice our test is nearly perfect, meaning that empirically, we never miss a true instance, and eliminate almost all false instances.

We note that while the idea of using distinguishing attacks to eliminate false instances in slide attacks was already known in the literature [2], to the best of our knowledge, nobody has ever explored this direction. Our analyses of FF3 thus provide a pedagogical example of this paradigm.

Contribution: A better attack on 4-round Feistel. Thanks to the LHD tests above, we are now left with $O(N)$ false instances and a few true instances. If one uses DV's codebook recovery attack on 4-round Feistel, one would end up with $O(N^4)$ expected time, which is still very expensive. The core part of DV's attack needs to find all directed 3-cycles of zero weight in a (random) directed graph $\mathcal{G} = (V, E)$. DV's approach is to enumerate all directed 3-cycles via some sparse matrix multiplications, and then pick those of zero weight, spending $O(|V| \cdot |E|)$ time. We instead give an elementary algorithm that uses $O(|V| + |E|)$ expected time. In addition, DV's attack relies on a conjecture of Feistel networks. They however can only empirically verify this conjecture for $N \in \{2, 2^2, \ldots, 2^9\}$. In this work, we resolve this conjecture, solving an open problem posed by DV.

Our algorithm above leads to the best known-plaintext attack to 4-round Feistel in the literature, using $O(N^{5/3})$ data and time complexity. A prior work by Biryukov, Leurent, and Perrin [9] is slightly better, recovering the codebook within $O(N^{3/2})$ data and time, but this attack requires chosen plaintexts and ciphertexts. A comparison of the attacks on 4-round Feistel is listed in Table 2.

Other contributions. We also generalize our FF3 attack to unbalanced settings, for a general domain $\mathbb{Z}_M \times \mathbb{Z}_N$, with $M \geq N \geq 64$, so that we can recover,

say encrypted SSNs. The asymmetry $M \geq N$ however requires some care in the extension of the attack on 4-round Feistel. In particular, due to the symmetry of 4-round Feistel, given plaintext/ciphertext pairs $(M_1, C_1), \ldots, (M_p, C_p)$, one can view $M_1, \ldots, M_p$ as the "ciphertexts" of $C_1, \ldots, C_p$ under an inverse 4-round Feistel, leading to a dual attack. The two attacks yield no difference in the balanced setting $M = N$, but if $M \gg N$, we find that the dual attack provides a superior recovery rate. We also introduce some tricks that substantially improve both the data complexity and the recovery rate.

On the other hand, there are often some gaps between the choices of the parameters according to DV's analyses, and what their experiments suggest. Even worse, the performance of their attacks is highly sensitive: in some experiments, if they triple the number of plaintext/ciphertext pairs, ironically, the recovery rate drops from 77% to 0%. DV thus have to calibrate concrete choices of the parameters via extensive experiments. In contrast, we choose to err on the conservative side in our analyses, and our estimates are consistent with the experiments. We also add some fail-safe to avoid the performance degradation when the number of plaintext/ciphertext pairs increases.

Limitation of our attack on FF3. Our attack exploits the same bug of FF3 as DV's attack, and thus it can be thwarted without hurting performance by restricting the tweak space, as DV suggested. In addition, both of our attack and DV's requires that the adversary can *adaptively* make chosen plaintexts on $\Theta(N^2)$ queries for domain $\mathbb{Z}_N \times \mathbb{Z}_N$, but it is unclear how to mount this kind of attack, especially with that many queries, in practice.

Additional related work. There have been two separate lines of building FPE schemes. On the theoretical side, we have provably secure constructions that are based on card shuffling, such as Swap-or-Not [18], Mix-and-Cut [26], or Sometimes-Recurse [22] that are too slow for performance-hungry applications. On the practical side, in addition to FF1/FF3, there are other industry proposals, such as FNR from Cisco [14], or DTP from Protegrity [21], that have no theoretical justification. Hoang, Tessaro, and Trieu [20] however show that FNR is somewhat vulnerable in tiny domains, and DTP is completely broken even in large domains.

In a different direction, Bellare and Hoang [3] study the security of DFF, an FPE scheme currently proposed to NIST for standardization [28], and show that for appropriately large domains, DFF provides a way to localize and limit the damage from key exposure. However, as DFF is based on a 10-round Feistel network, it is still subject to prior attacks on generic Feistel-based FPE [5, 20] on tiny domains.

Very recently, Durak and Vaudenay [16] give some theoretical codebook-recovery attacks on generic balanced r-round Feistel, for $r \geq 5$. They conclude that on domain $\mathbb{Z}_N \times \mathbb{Z}_N$, FF1 cannot provide 128-bit security for $N \leq 11$, and FF3 for $N \leq 17$.

2 Preliminaries

NOTATION. If y is a string then let $|y|$ denote its length and let $y[i]$ denote its i-th bit for $1 \le i \le |y|$. We write $y[i:j]$ to denote the substring of y, from the ith bit to the j-th bit, inclusive. 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. We use the code based game playing framework of [7]. In particular, by $\Pr[\mathrm{G}]$ we denote the probability that the execution of game G returns true.

FPE. An FPE scheme F is a pair of deterministic algorithms $(\mathsf{F.E}, \mathsf{F.D})$, where $\mathsf{F.E} : \mathsf{F.Keys} \times \mathsf{F.Twk} \times \mathsf{F.Dom} \to \mathsf{F.Dom}$ is the encryption algorithm, $\mathsf{F.D} : \mathsf{F.Keys} \times \mathsf{F.Twk} \times \mathsf{F.Dom} \to \mathsf{F.Dom}$ the decryption algorithm, $\mathsf{F.Keys}$ the key space, $\mathsf{F.Twk}$ the tweak space, and $\mathsf{F.Dom}$ the domain. For every key $K \in \mathsf{F.Keys}$ and tweak $T \in \mathsf{T}$, the map $\mathsf{F.E}(K, T, \cdot)$ is a permutation over $\mathsf{F.Dom}$, and $\mathsf{F.D}(K, T, \cdot)$ reverses $\mathsf{F.E}(K, T, \cdot)$.

FEISTEL-BASED FPES. Most existing FPE schemes, including FF3, are based on Feistel networks. Following BHT [5], we specify Feistel-based FPE in a general, parameterized way. This allows us to refer to both schemes of ideal round functions for the analysis, and schemes of some concrete round functions for realizing 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 $r \ge 2$ is an integer, the number of rounds, and $\boxplus$ is 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. Integers $M, N \ge 1$ define the domain of F as $\mathsf{F.Dom} = \mathbb{Z}_M \times \mathbb{Z}_N$. The parameter $\mathrm{PL} = (\mathcal{T}, \mathcal{K}, F_1, \ldots, F_r)$ specifies the set $\mathcal{T}$ of tweaks and a set $\mathcal{K}$ of keys, meaning $\mathsf{F.Twk} = \mathcal{T}$ and $\mathsf{F.Keys} = \mathcal{K}$, and the round functions $F_1, \ldots, F_r$ such that $F_i : \mathcal{K} \times \mathcal{T} \times \mathbb{Z}_N \to \mathbb{Z}_M$ 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 code of $\mathsf{F.E}$ and $\mathsf{F.D}$ is shown in Fig. 1.

Classic Feistel schemes correspond to the boolean case, where $M = 2^m$ and $N = 2^n$ are powers of two, and $\boxplus$ is the bitwise xor operator $\oplus$. The scheme is balanced if $M = N$ and unbalanced otherwise. 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. We write $\mathsf{LH}(X)$ and $\mathsf{RH}(X)$ to refer to the left and right segments of X respectively. For simplicity, we assume that 0 is the zero element of the groups $(\mathbb{Z}_M, \boxplus)$ and $(\mathbb{Z}_N, \boxplus)$.

FEISTEL-BASED BLOCKCIPHERS. If the tweak space $\mathcal{T}$ is a singleton set then FPE degenerates into a blockcipher (of a general domain). For such a blockcipher F, we write $\mathsf{F.E}(K, M)$ and $\mathsf{F.D}(K, C)$ instead of $\mathsf{F.E}(K, T, M)$ and $\mathsf{F.D}(K, T, C)$ respectively.

In our analysis of Feistel-based blockciphers, the round functions are modeled as truly random. We write $\mathbf{Feistel}[r, M, N, \boxplus]$ to denote $\mathbf{Feistel}[r, M, N, \boxplus, \mathrm{PL}]$, for the ideal choice of $\mathrm{PL} = (\mathcal{T}, \mathcal{K}, F_1, \ldots, F_r)$ in which (i) $\mathcal{T} = \{\varepsilon\}$ where ε is the empty string, and (ii) $\mathcal{K}$ is the set $\mathbf{RF}(r, M, N)$ of all tuples of functions $(G_1, \ldots, G_r)$ such that $G_i : \mathbb{Z}_N \to \mathbb{Z}_M$ if i is odd, and $G_i : \mathbb{Z}_M \to \mathbb{Z}_N$ if i is

$\underline{\mathsf{F.E}(K,T,X)}$
$(L,R) \leftarrow X$
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)

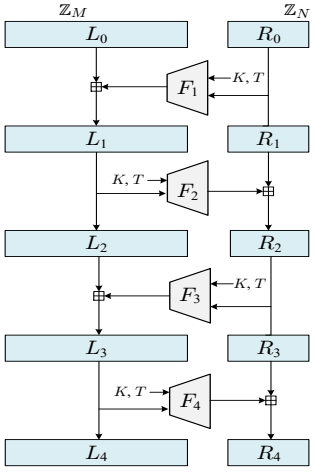

Fig. 1. Left: The code for the encryption and decryption algorithms of $\mathsf{F} = \textbf{Feistel}[r, M, N, \boxplus, \mathrm{PL}]$ where $\mathrm{PL} = (\mathcal{T}, \mathcal{K}, F_1, \ldots, F_r)$. **Right:** An illustration of encryption with $r = 4$ rounds.

even, and (iii) for $1 \le i \le r$, the function $F_i(K, \cdot)$ is defined as $G_i(\cdot)$, where $(G_1, \ldots, G_r) \leftarrow K$.

3 Breaking FF3

In this section, we describe a chosen-plaintext codebook-recovery attack on FF3 that we call *Slide-then-Differential* (SD) attack.[4] This is based on Triangle-Finding (TF) attack, a known-plaintext codebook-recovery attack on 4-round Feistel that we will present in the next section. The running time of TF is $O(M^{5/3})$, and it actually recovers the round functions of the Feistel network, using

$$p = \max\Big\{ \lfloor 2^{1/3} M^{2/3} N \rfloor, \lceil M(\ln(M) + 5) \rceil \Big\} \tag{1}$$

known plaintext/ciphertext pairs. We note that TF is used in a modular way; one does not need to know its technical details to understand SD.

The FF3 scheme. FF3 is a Feistel-based FPE scheme $\mathsf{F} = \textbf{Feistel}[8, M, N, \boxplus, \mathrm{PL}]$ of 8 rounds, where M and N are integers such that $M \ge N \ge 2$ and $MN \ge 100$.[5] The parameter PL specifies tweak space $\mathsf{F.Twk} = \{0,1\}^{2\tau}$, and

[4] While the notion of chosen-plaintext codebook-recovery attacks on blockciphers is folklore, one has to exercise some care in carrying this notion to FPE, because FPE domains can be tiny. In Appendix A, we give a formal definition of chosen-plaintext codebook-recovery attacks on FPE.

[5] In NIST specification, the $\boxplus$ operation is the modular addition in $\mathbb{Z}_N$ and $\mathbb{Z}_M$, but here we will consider a generic group operator. Moreover, FF3 uses near-balanced

two keyed hash functions $H_1 : \mathsf{F.Keys} \times \{0,1\}^\tau \times \mathbb{Z}_N \to \mathbb{Z}_M$, and $H_2 : \mathsf{F.Keys} \times \{0,1\}^\tau \times Z_M \to \mathbb{Z}_N$. For each $i \le 8$, if i is odd then the round function F_i is constructed via $F_i(K,T,X) = H_1(K, T[1:\tau] \oplus [i-1]_\tau, X)$, otherwise if i is even then $F_i(K,T,X) = H_2(K, T[\tau+1:2\tau] \oplus [i-1]_\tau, X)$, where $[j]_\tau$ is a τ-bit encoding of the integer j and $\oplus$ is the bitwise xor operator.

In analysis, the hash functions H_1 and H_2 are modeled as truly random. Formally, let $\mathcal{K}$ be the set $\mathbf{RF}(\tau, M, N)$ of all pairs of functions (G_1, G_2) such that $G_1 : \{0,1\}^{2\tau} \times \mathbb{Z}_N \to \mathbb{Z}_M$, and $G_2 : \{0,1\}^{2\tau} \times \mathbb{Z}_M \to \mathbb{Z}_N$. Then for each $j \le 2$, define $H_j(K,\cdot,\cdot) = G_j(\cdot,\cdot)$, where $(G_1, G_2) \leftarrow K$, and we write $\mathbf{FF3}[M, N, \tau, \boxplus]$ to denote this ideal version of FF3.

In our attack to FF3, we will consider $M \ge N \ge 64$ and $MN \ge 2p$, where p is specified as in Equation (1). While there are indeed applications of smaller values of M and N, they are already susceptible to prior attacks [15, 4, 20] whose running time is practical in those tiny domains. In addition, to simplify our asymptotic analysis, we will assume that $N = \Omega(\sqrt{M})$, which applies to the setting of the FF3 scheme, since FF3 uses near-balanced Feistel. Thus $p \in O(M^{2/3}N)$.

3.1 DV's Blueprint for Breaking FF3

Let $\mathsf{F} = \mathbf{FF3}[M, N, \tau, \boxplus]$. Let K and T be a key and tweak for F, respectively. Recall that $\mathsf{F.E}(K, T, \cdot)$ is an 8-round Feistel network. View $\mathsf{F.E}(K, T, \cdot)$ as the cascade of two 4-round Feistel networks f and g, meaning $\mathsf{F.E}(K, T, X) = g(f(X))$ for every $X \in \mathbb{Z}_M \times \mathbb{Z}_N$. DV [15] observe that $\mathsf{F.E}(K, T', \cdot)$ is the cascade of g and f—note that the ordering of f and g is now reversed—where $T' = T \oplus ([4]_\tau \,\|\, [4]_\tau)$. See Fig. 2 for an illustration.

A sketch of DV's attack. From the observation above, one can launch a chosen-plaintext codebook-recovery attack on F as follows; this can also be viewed as a slide attack [10, 11]. Let p be as specified in Equation (1) and let $s = \left\lfloor \sqrt{MN/2p} \right\rfloor \ge 1$.[6] Sample s elements uniformly and independently from $\mathbb{Z}_M \times \mathbb{Z}_N$, and let S be the set of these elements. Repeat this process, and let S^* be the resulting set. Recall that the adversary is given an encryption oracle Enc in this attack. Now, for each $U_0 \in S$, we iterate $U_i \leftarrow \textsc{Enc}(T, U_{i-1})$, for $i = 1, \ldots, 2p$, forming a $\mathtt{U}$-*chain* $U_0 \to U_1 \to \cdots \to U_{2p}$. For each $V_0 \in S^*$, let $V_i \leftarrow \textsc{Enc}(T', V_{i-1})$ for $i = 1, \ldots, 2p$, forming a $\mathtt{V}$-*chain* $V_0 \to V_1 \to \cdots \to V_{2p}$.

Consider a $\mathtt{U}$-chain and a $\mathtt{V}$-chain such that each chain has at least p distinct elements.[7] If there is some index $i < p$ such that $V_0 = f(U_i)$ then the pair (U_i, V_0) is called a *slid pair*, and $V_k = f(U_{i+k})$ and $U_{i+k+1} = g(V_k)$ for every

Feistel, and thus the values of M and N are very close: if one wants to encrypt m characters in radix d, then $M = d^{\lceil m/2 \rceil}$ and $N = d^{\lfloor m/2 \rfloor}$.

[6] DV actually use different concrete choices of p and s to aggressively improve the recovery rate.

[7] To test if, say a $\mathtt{U}$-chain $(U_0, \ldots, U_{2p})$ contains at least p distinct elements, we only need to check if $U_0 \notin \{U_1, \ldots, U_{p-1}\}$, since $|\{U_0, \ldots, U_{2p}\}| < p$ if and only if U_0 is within a cycle of length $k < p$ in the functional graph of the permutation $f(g(\cdot))$.

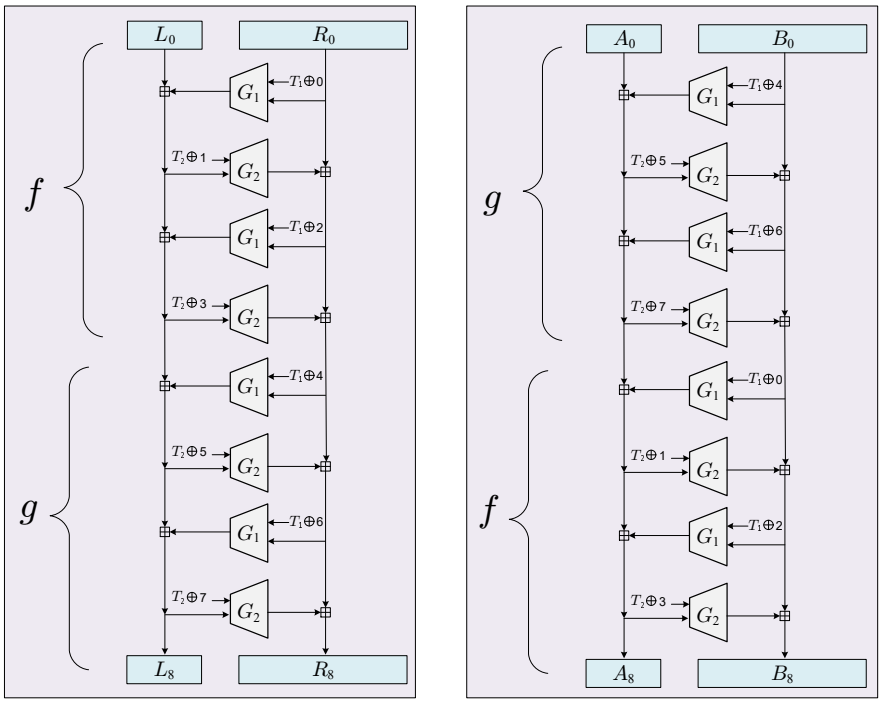

Fig. 2. Left: Encryption $\mathsf{F.E}(K, T, \cdot)$ as a cascade of 4-round Feistel networks f and g. **Right:** Slided encryption $\mathsf{F.E}(K, T', \cdot)$ as a cascade of g and f, with $T' = T \oplus ([4]_\tau \,\|\, [4]_\tau)$. Here T_1 and T_2 are the left half and right half of the tweak T, respectively. For simplicity, in the picture, instead of writing, say $T_1 \oplus [0]_\tau$, we simply write $T_1 \oplus 0$.

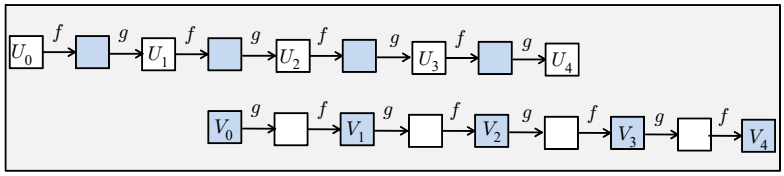

Fig. 3. Illustration of the slide attack. Here (U_1, V_0) is a slid pair.

$0 \le k < p$. Likewise, if there is some index $j < p$ such that $U_0 = g(V_j)$ then the pair (V_j, U_0) is also called a slid pair, and $U_k = g(V_{j+k})$ and $V_{j+k+1} = f(U_k)$ for every $0 \le k < p$. See Fig. 3 for an illustration.

Suppose that somehow we manage to find a slid pair. Then we get p input/output pairs for f, and can run TF to recover the codebook of f. Likewise, we can also recover the codebook of g. By composing the codebook of f and g, we finally recover the codebook of F on tweak T. We can also compose g and f to recover the codebook of F on tweak T'.

Procedure $\mathsf{SD}^{\textsc{Enc}}()$

Pick arbitrary $T \in \{0,1\}^{2\tau}$; $T' \leftarrow T \oplus ([4]_\tau \,\|\, [4]_\tau)$
$UCh \leftarrow \mathsf{MakeChain}^{\textsc{Enc}}(T)$; $VCh \leftarrow \mathsf{MakeChain}^{\textsc{Enc}}(T')$
For $\boldsymbol{U} \in UCh, \boldsymbol{V} \in VCh$ do
  $\boldsymbol{C} \xleftarrow{\$} \mathsf{Slide}(\boldsymbol{U}, \boldsymbol{V})$; If $\boldsymbol{C} \neq \perp$ then return $(T, \boldsymbol{C})$
  $\boldsymbol{C}' \xleftarrow{\$} \mathsf{Slide}(\boldsymbol{V}, \boldsymbol{U})$; If $\boldsymbol{C}' \neq \perp$ then return $(T', \boldsymbol{C}')$

Procedure $\mathsf{MakeChain}^{\textsc{Enc}}(T)$

$p \leftarrow \max\left\{\lfloor 2^{1/3} M^{2/3} N \rfloor, \lceil M(\ln(M)+5) \rceil\right\}$
$s \leftarrow \lfloor \sqrt{MN/2p} \rfloor$; $S, UCh \leftarrow \emptyset$
For $i = 1$ to s do $U \xleftarrow{\$} \mathbb{Z}_M \times \mathbb{Z}_N$; $S \leftarrow S \cup \{U\}$
For $U_0 \in S$ do
  For $i = 1$ to $2p$ do $U_i \leftarrow \textsc{Enc}(T, U_{i-1})$
  If $U_0 \notin \{U_1, \ldots, U_{p-1}\}$ then $UCh \leftarrow UCh \cup \{(U_0, \ldots, U_{2p})\}$
Return UCh

Fig. 4. The blueprint of DV's attack, which is also the main procedure of the SD attack. Procedure $\mathsf{Slide}(\boldsymbol{U}, \boldsymbol{V})$ takes as input two chains $\boldsymbol{U} = (U_0, \ldots, U_{2p})$ and $\boldsymbol{V} = (V_0, \ldots, V_{2p})$, tries to find a slid pair (U_i, V_0), and then uses TF to recover the codebook. Numbers M, N, τ are global parameters.

The code of the blueprint of DV's attack is given in Fig. 4, which is also the main procedure of our SD attack. The two attacks however differ in how they implement procedure Slide for finding a slid pair, among $2s^2p \approx MN$ candidates. DV simply try every possible candidate, by running (a slow version of) the TF algorithm to recover the codebook of f and g. As we will show below, there are often very few slid pairs, and thus DV's attack essentially has to run TF for about $\Theta(MN)$ times, which is very expensive. The key idea in our Slide-then-Differential, which we will elaborate in Section 3.2, is to use some differential analysis to quickly eliminate false candidates.

The number of slid pairs. Clearly, the attack above only works if there exists at least one slid pair. Let P be the random variable for the number of slid pairs. DV use a heuristic[8] to estimate that $\Pr[P \geq 1] \approx 1 - e^{-2s^2p/MN} \approx 1 - 1/e$, under the model that the cascade of f and g is an ideal permutation. We instead give a rigorous lower bound of $\Pr[P \geq 1]$ for a *generic* value p in Lemma 1 in the same model; the proof is in Appendix C.1. For $s = 1$ (equivalently, $M < 1024$), we can compute the exact probability $\Pr[P \geq 1]$, but we stress that this result only holds in the model above. The experiments in Section 5 show that empirically, the event $P \geq 1$ happens with higher probability.

[8] While DV only consider balanced Feistel networks, their heuristic can be easily generalized to the general case. For completeness, in the proof of Lemma 1, we also describe this heuristic argument.

Lemma 1. *Let $M \geq N \geq 8$ and let $\mathsf{F} = \mathbf{FF3}[M, N, \tau, \boxplus]$. Let $p \geq 1$ be an integer such that $p \leq MN/2$ and let $s = \lfloor \sqrt{MN/2p} \rfloor$. Let f and g be as above, and let π be the cascade of f and g. Let P be the random variable for the number of slid pairs. Let $\delta = \frac{2p}{MN} - \frac{(2.5p^2 - 1.5p)}{(MN)^2}$. We will model π as an ideal random permutation on $\mathbb{Z}_M \times \mathbb{Z}_N$.*

(a) If $s = 1$ then $\Pr[P \geq 1] = \delta \approx \frac{3}{8}$.
(b) If $s \geq 2$ then $\Pr[P \geq 1] \geq \frac{s^2\delta}{2} \approx \frac{1}{2}$.

Above, we show that it is quite likely that there are one or more slid pairs. However, often there will be very few of them. Lemma 2 below bounds the expected number of slid pairs for a *generic* value of p; the proof is in Appendix C.2. Combining this with Markov's inequality, one can show that with probability at least 0.8, there are at most 5 slid pairs.

Lemma 2. *Let $M \geq N \geq 64$ and let $\mathsf{F} = \mathbf{FF3}[M, N, \tau, \boxplus]$. Let $p \geq 1$ be an integer such that $p \leq MN/2$ and let $s = \lfloor \sqrt{MN/2p} \rfloor$. Let f and g be as above, and let π be the cascade of f and g. Let P be the random variable for the number of slid pairs. If we model π as an ideal random permutation on $\mathbb{Z}_M \times \mathbb{Z}_N$ then $\mathbf{E}[P] \leq \frac{2s^2p}{MN} \leq 1$.*

3.2 Distinguishing Slid-pair Candidates

As shown above, we often have very few slid pairs, among $2s^2p \approx MN$ candidates, and using TF to find the actual slid pairs is an overkill. Note that each candidate gives us p plaintext/ciphertext pairs for f. If a candidate is indeed a slid pair, then the ciphertexts for f are indeed the images of the corresponding plaintexts under f, otherwise we can view them as produced from an ideal permutation on $\mathbb{Z}_M \times \mathbb{Z}_N$. The analogous claim also holds for g. A natural solution is to find a quick distinguishing attack for 4-round Feistel, so that we can tell the true candidates from the false ones.

Our distinguishing attack on 4-round Feistel. Below, we will give a distinguishing attack Left-Half Differential (LHD) of 4-round Feistel such that (1) in the ideal world, it returns 1 with probability at most $\frac{N^{5/6}}{M^{4/3}}$, and (2) in the real world, it returns 1 with probability at least $1 - \frac{\sqrt{N}}{8M} - \frac{9.7}{M} - \frac{0.88N^{3/4}}{M^{3/2}}$. For each slid-pair candidate, we run LHD on the plaintext/ciphertext pairs of f, and also on those of g. We will accept the candidate if LHD returns 1 in both cases. Then, for each false candidate, the chance that we incorrectly accept it is at most $N^{5/3}/M^{8/3}$. Since we have at most MN false candidates, on average we will have at most

$$MN \cdot \frac{N^{5/3}}{M^{8/3}} = \frac{N^{8/3}}{M^{5/3}} \leq N$$

false candidates that survived our test. In addition, for each true candidate, the chance that we incorrectly reject it is at most

$$1 - \left(1 - \frac{\sqrt{N}}{8M} - \frac{9.7}{M} - \frac{0.88N^{3/4}}{M^{3/2}}\right)^2 \leq 0.37$$

for $M, N \geq 64$. We note that our bounds are very conservative, since we obtain them via Chebyshev's inequality, which is loose. In fact, our empirical results, presented in Section 5, significantly outperform the theoretical estimates. In particular, on average just one (possibly false) candidate survives our test, and we never incorrectly reject a true candidate.

Proceeding into details, the LHD algorithm is based on the following Lemma 3, which is a generalization of a result by Patarin for balanced, boolean Feistel [24]. A more general version of Lemma 3 appears in [5] for a Feistel network of an even number of rounds, but this result only provides a (very tight) approximation of the bound, instead of an exact one.

Lemma 3. *Let $M, N \geq 8$ be integers and $\overline{\mathsf{F}} = \mathbf{Feistel}[4, M, N, \boxplus]$. Let X and X' be two distinct messages in $\mathbb{Z}_M \times \mathbb{Z}_N$ such that $\mathsf{RH}(X) = \mathsf{RH}(X')$. Let C and C' be the ciphertexts of X and X' under $\overline{\mathsf{F}}$ with a uniformly random key. Then*

$$\Pr[\mathsf{LH}(C) \boxminus \mathsf{LH}(C') = \mathsf{LH}(X) \boxminus \mathsf{LH}(X')] = \frac{M + N - 1}{MN} \ .$$

The proof of Lemma 3 is in Appendix C.3. It is based on the following technical result that is also needed in several other places; the proof is in Appendix C.4.

Lemma 4. *Let $M, N \geq 8$ be integers and $\overline{\mathsf{F}} = \mathbf{Feistel}[4, M, N, \boxplus]$. Let X and X' be two distinct messages in $\mathbb{Z}_M \times \mathbb{Z}_N$ such that $\mathsf{RH}(X) = \mathsf{RH}(X')$. Let C and C' be the ciphertexts of X and X' under $\overline{\mathsf{F}}$ with a uniformly random key. Let X_t and X'_t be the round-t intermediate outputs of X and X' respectively.*

(a) The random variables $\mathsf{RH}(X_2)$ and $\mathsf{RH}(X'_2)$ are uniformly and independently distributed over $\mathbb{Z}_N$.
(b) If $\mathsf{RH}(X_2) = \mathsf{RH}(X'_2)$ then $\mathsf{LH}(C) \boxminus \mathsf{LH}(C') = \mathsf{LH}(X) \boxminus \mathsf{LH}(X')$ with certainty.
(c) Fix distinct $R, R' \in \mathbb{Z}_N$. If $\mathsf{RH}(X_2) = R$ and $\mathsf{RH}(X'_2) = R'$ then $\mathsf{LH}(C) \boxminus \mathsf{LH}(C') = \mathsf{LH}(X) \boxminus \mathsf{LH}(X')$ with probability $1/M$.

Lemma 3 above shows that if we encrypt two messages X and X' of the same right segment under a 4-round Feistel network, then there will be some bias in the distribution of the ciphertexts C and C': (1) the chance that $\mathsf{LH}(C) \boxminus \mathsf{LH}(C') = \mathsf{LH}(X) \boxminus \mathsf{LH}(X')$ is $\frac{M+N-1}{MN}$, (2) had we instead sampled C and C' uniformly without replacement from $\mathbb{Z}_M \times \mathbb{Z}_N$, this probability would have been just $\frac{N}{MN-1}$. Our distinguishing attack LHD will amplify this bias, by using several messages of the same right segments.

Random variables $X_1, \ldots, X_m \in \mathbb{Z}_M \times \mathbb{Z}_N$ are *t-wise right-matching* if they satisfy the following constraints:

- If we partition $X_1, \ldots, X_m$ into groups $P_1, \ldots, P_d$ according to their right segments then $d \leq t$.
- Within each partition P_i, the left segments of the messages in P_i are uniformly distributed over $\mathbb{Z}_M$, subject to the constraint that those left segments are distinct.

Procedure $\mathsf{LHD}(X_1, \ldots, X_m, C_1, \ldots, C_m)$
Partition $X_1, \ldots, X_m$ by the right segments into groups $P_1, \ldots, P_d$
$\mathit{count} \leftarrow 0$; $\Delta \leftarrow \frac{1}{5} \cdot \frac{M+N-1}{MN} + \frac{4}{5} \cdot \frac{N}{MN-1}$; $\mathit{size} \leftarrow \sum_{\ell=1}^{d} \frac{|P_\ell|(|P_\ell|-1)}{2}$
For $\ell \leftarrow 1$ to d do
  For $X_i, X_j \in P_\ell$ with $i < j$ do
    If $\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j) = \mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j)$ then $\mathit{count} \leftarrow \mathit{count} + 1$
If $\mathit{count} \geq \Delta \cdot \mathit{size}$ then return 1 else return 0

Fig. 5. Distinguishing attack LHD on four-round Feistel.

Our attack LHD takes as input m messages $(X_1, \ldots, X_m)$ that are t-wise right-matching and their ciphertexts $(C_1, \ldots, C_m)$, where $m = \left\lceil \frac{p}{N} \cdot \lceil 32N^{1/6} \rceil \right\rceil$ and $t = \lceil \frac{mM}{p} \rceil$. The code of LHD is given in Fig. 5. Informally, LHD will compute *count*, the number of pairs X_i and X_j, with $i < j$, such that $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$ and $\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j) = \mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j)$. If the ciphertexts are produced by a 4-round Feistel network then from Lemma 3, the expected value of *count* is $\frac{M+N-1}{MN} \cdot \mathit{size}$, where *size* is the number of pairs X_i, X_j such that $i < j$ and $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$. If the ciphertexts are produced by a truly random permutation on $\mathbb{Z}_M \times \mathbb{Z}_N$ then the expected value of *count* is $\frac{N}{MN-1} \cdot \mathit{size}$. The algorithm LHD will return 1 if *count* is greater than the weighted average $\left(\frac{1}{5} \cdot \frac{M+N-1}{MN} + \frac{4}{5} \cdot \frac{N}{MN-1}\right)\mathit{size}$, otherwise it will return 0.

Implementing LHD. The code in Fig. 5 describes just the conceptual view of LHD for ease of understanding. Implementing it efficiently requires some care. First, messages $X_1, \ldots, X_m$ will be grouped according to their right segments, by a one-time preprocessing that we will describe in Section 3.3. Thus the partitioning takes only linear time. Let $P_1, \ldots, P_d$ be the resulting partitions, and let $|m_\ell| = |P_\ell|$, for every $\ell \leq d$. In the for loops, if we naively follow the code, then the running time would be

$$\sum_{\ell=1}^{d} \Omega(m_\ell^2) = \Omega(m^2/d) = \Omega(M^{1/3}N^{7/6}),$$

which is expensive. Instead, we will execute as in Fig. 6. That is,

- For each fixed $\ell \leq d$, we want to find count_ℓ, the number of pairs (i, j) such that $i < j$ and $X_i, X_j \in P_\ell$ and $\mathsf{LH}(C_i) \boxminus \mathsf{LH}(X_i) = \mathsf{LH}(C_j) \boxminus \mathsf{LH}(X_j)$. We then can compute *count* via $\mathit{count}_1 + \cdots + \mathit{count}_d$.
- Thus for each $\ell \leq d$, we create an empty hash table H_ℓ of key-value pairs and initialize $\mathit{count}_\ell \leftarrow 0$. We process P_ℓ so that eventually, for each entry in H_ℓ, its key is a number $Z \in \mathbb{Z}_M$ and its value indicates how many $X_k \in P_\ell$ that $\mathsf{LH}(C_k) \boxminus \mathsf{LH}(X_k) = Z$.
- Finally, we iterate through all keys of H_ℓ. For each key Z, we find its value v and update $\mathit{count}_\ell \leftarrow \mathit{count}_\ell + \frac{v(v-1)}{2}$.

The total running time of this implementation is $O(m) = O(M^{2/3}N^{1/6})$.

Procedure $\mathsf{LHD}(X_1, \ldots, X_m, C_1, \ldots, C_m)$

// $X_1, \ldots, X_m$ are already grouped according to their right segments
Partition $X_1, \ldots, X_m$ by the right segments into groups $P_1, \ldots, P_d$
$\mathit{count} \leftarrow 0$; $\Delta \leftarrow \frac{1}{5} \cdot \frac{M+N-1}{MN} + \frac{4}{5} \cdot \frac{N}{MN-1}$; $\mathit{size} \leftarrow \sum_{\ell=1}^{d} \frac{|P_\ell|(|P_\ell|-1)}{2}$
For $\ell \leftarrow 1$ to d do
  $\mathit{count}_\ell \leftarrow 0$; Initialize a hash table H_ℓ
  For $X_k \in P_\ell$ do
    $Z \leftarrow \mathsf{LH}(C_k) \boxminus \mathsf{LH}(X_k)$; $v \leftarrow H_\ell[Z]$
    If $v = \perp$ then $H_\ell[Z] \leftarrow 1$ else $H_\ell[Z] \leftarrow v + 1$
  For each key Z in H_ℓ do $v \leftarrow H_\ell[Z]$; $\mathit{count}_\ell \leftarrow \mathit{count}_\ell + \frac{v(v-1)}{2}$
$\mathit{count} \leftarrow \mathit{count}_1 + \cdots + \mathit{count}_d$
If $\mathit{count} \geq \Delta \cdot \mathit{size}$ then return 1 else return 0

Fig. 6. Implementation of LHD.

Analysis of LHD. Lemma 5 below bounds the probability that LHD outputs 1 in the ideal world, for *generic* m and t, and also for a *generic* weighted average $\Delta = \lambda \cdot \frac{M+N-1}{MN} + (1-\lambda)\frac{N}{MN-1}$; see Appendix C.5 for the proof. If we pick $m = \left\lceil \frac{p}{N} \cdot \lceil 32N^{1/6} \rceil \right\rceil$, $t = \lceil \frac{mM}{p} \rceil$, and $\lambda = \frac{1}{5}$ as suggested then this probability is about $\frac{N^{5/6}}{M^{4/3}}$.

Lemma 5. *Let $M \geq N \geq 8$ be integers, and let $0 < \lambda < 1$ be a real number. Let $m > t \geq 1$ be integers. Let $X_1, \ldots, X_m$ be t-wise right-matching messages, and let $C_1, \ldots, C_m$ be their ciphertexts, respectively, under an ideal random permutation on $\mathbb{Z}_M \times \mathbb{Z}_N$. Let V be the random variable of the number of pairs X_i and X_j, with $i < j$, such that $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$ and $\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j) = \mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j)$. Let size be the number of pairs X_i, X_j such that $i < j$ and $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$, and $\Delta = \lambda \cdot \frac{M+N-1}{MN} + (1-\lambda)\frac{N}{MN-1}$. Then*

$$\Pr[V \geq \Delta \cdot \mathit{size}] \leq \frac{N^2}{\lambda^2(M-2)^2}\left(\frac{1}{MN-2} + \frac{2MN-2}{N(m^2/t-m)}\right) .$$

Lemma 6 below bounds the probability that LHD fails to output 1 in the real world, again for generic m and t, and for a generic weighted average $\Delta = \lambda \cdot \frac{M+N-1}{MN} + (1-\lambda)\frac{N}{MN-1}$; see Appendix C.6 for the proof. If we use $m = \left\lceil \frac{p}{N} \cdot \lceil 32N^{1/6} \rceil \right\rceil$, $t = \lceil \frac{mM}{p} \rceil$, and $\lambda = \frac{1}{5}$ as suggested then this probability is about $\frac{\sqrt{N}}{8M} + \frac{9.7}{M} + \frac{0.88N^{3/4}}{M^{3/2}}$.

Lemma 6. *Let $M \geq N \geq 8$ be integers and let $0 < \lambda < 1$ be a real number. Let $m > t \geq 1$ be integers. Let $X_1, \ldots, X_m$ be t-wise right-matching messages and let $C_1, \ldots, C_m$ be their ciphertexts, respectively, under $\overline{\mathsf{F}} = \mathbf{Feistel}[4, M, N, \boxplus]$ with a uniformly random key. Let V be the random variable of the number of pairs X_i and X_j, with $i < j$, such that $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$ and $\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j) =$*

$\mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j)$. *Let* $\Delta = \lambda \cdot \frac{M+N-1}{MN} + (1-\lambda)\frac{N}{MN-1}$, *and let size be the number of pairs* X_i, X_j *such that* $i < j$ *and* $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$. *Then*

$$\Pr\left[V \le \Delta \cdot size\right] \le \frac{2(M+N-1)MN}{(1-\lambda)^2(m^2/t-m)(M-2)^2} + \frac{6.2(M-1)(N-1)}{(1-\lambda)^2 N(M-2)^2} + \frac{4MN}{(1-\lambda)^2(M-2)^2\sqrt{(m^2/t-m)}} \ .$$

Using LHD. The LHD attack requires m chosen plaintexts, but recall that for each slid-pair candidate, we only have p known plaintext/ciphertext pairs for f, and also p known pairs for g.[9] To find m messages that are $\lceil \frac{mM}{p} \rceil$-wise right-matching, the naive approach is to partition p given messages according to their right segments, and let $P_1, \ldots, P_M$ be the (possibly empty) partitions, with $|P_1| \ge \cdots \ge |P_M|$. We then output m messages from the first (and also biggest) $s = \left\lceil \frac{mM}{p} \right\rceil$ partitions. Since there are at most M partitions, our chosen partitions contain at least $\lceil s \cdot \frac{p}{M} \rceil \ge m$ messages. Moreover, as the given p messages are sampled uniformly without replacement from $\mathbb{Z}_M \times \mathbb{Z}_N$, within each partition P_i, the left segments of the messages in P_i are sampled uniformly without replacement from $\mathbb{Z}_M$.

The naive approach above is however very expensive. Totally, for $\Theta(MN)$ slid-pair candidates, it uses $\Omega(MNp) = \Omega(M^{5/3}N^2)$ time just to find their right-matching messages. In the next section we'll describe a one-time preprocessing of $O(MN)$ time such that later for each slid-pair candidate, we need only $O(m)$ time to find their right-matching messages to run LHD tests. Only for candidates that survive the LHD tests that we extract their p plaintext/ciphertext pairs from the corresponding chains to run TF.

Eliminating false negatives. Since our distinguishing test of slid-pair candidates above might occasionally produce false negatives, we still have to use TF to eliminate the survived false candidates. The TF algorithm, after recovering the round functions (G_1, G_2, G_3, G_4), will compute the outputs of the first $3M \le p$ plaintexts under a 4-round Feistel network with the round functions (G_1, G_2, G_3, G_4), and compare them with the corresponding ciphertexts. By a simple counting argument, one can show that it is extremely likely that TF will reject all these false candidates. Specifically, for a false candidate, view its $3M$ associated ciphertexts as the outputs of the $3M$ plaintexts under an ideal permutation on $\mathbb{Z}_M \times \mathbb{Z}_N$. On the one hand, there are at most $M^{2N}N^{2M} \le M^{2M}N^{2N}$ choices of four round functions for a 4-round Feistel network on $\mathbb{Z}_M \times \mathbb{Z}_N$. On the other hand, if we sample $3M$ ciphertexts uniformly without replacement from $\mathbb{Z}_M \times \mathbb{Z}_N$, there are

$$MN \cdots (MN - 3M + 1) \ge (MN - 3M)^{3M} \ge M^{3M}(N-3)^{3N}$$

[9] Recall that in our attack, we require $M \ge N \ge 64$. This ensures that $m \le p$, so that we can select m right-matching messages from p known messages.

equally likely outputs. Since we have to deal with at most MN false candidates, the chance that the TF algorithm fails to eliminate all false candidates is at most

$$\frac{MN \cdot M^{2M} N^{2N}}{M^{3M}(N-3)^{M+2N}} = \frac{N^{2N+1}}{M^{3M-1}(N-3)^{3N}} \leq \frac{1}{M^{3M-1}} \ .$$

Relation to prior Feistel attacks. Our LHD attack generalizes Patarin's distinguishing attack on 4-round balanced, boolean Feistel [24]; Patarin's result was later rediscovered by Aiello and Venkatesan [1]. To attack Feistel networks on $2n$-bit strings, those papers suggest using messages $X_1, \ldots, X_m$ of the same right half, with $m = \Theta(2^{n/2})$. However, both papers only compute the expected value of the number of pairs (i, j) such that $1 \leq i < j \leq m$ and $\mathsf{LH}(C_i) \oplus \mathsf{LH}(C_j) = \mathsf{LH}(X_i) \oplus \mathsf{LH}(X_j)$ in both the real and ideal worlds. Consequently, they cannot analyze the advantage of their attack, and can only suggest an asymptotic value of m. Our Lemma 5 and Lemma 6 allow one to fill this gap. By using $N = M = 2^n$, $t = 1$, $m = c \cdot 2^{n/2}$, and $\lambda = \frac{1}{2}$, the attack in [24, 1] achieves advantage around $1 - \frac{24}{c^2} - \frac{29}{2^n} - \frac{16}{c \cdot 2^{n/2}}$. Thus to achieve advantage $1/2$, for $n \geq 16$, we can use $c = 7$.

The attack in [24, 1] however cannot be used in place of LHD. Recall that we want a distinguishing attack that outputs 1 with probability around $1/\sqrt{N}$ or smaller in the ideal world, so that it can be used to eliminate most false slid-pair candidates. Using $m = \Theta(\sqrt{N})$ messages as suggested in [24, 1] does not meet this requirement, as the attack will output 1 with constant probability in the ideal world, according to Lemma 5.

3.3 Slide-then-Differential Attack

In this Section, we describe how to combine DV's slide attack with the LHD attack above, resulting in our Slide-then-Differential attack.

Speeding up with preprocessing. Recall that we have $\Theta(MN)$ slid-pair candidates, and for each such candidate we have to process $\Theta(p)$ pairs of plaintext/ciphertext. At the first glance it seems that we are doomed with $\Omega(pMN) = \Omega(M^{5/3}N^2)$ time. However, we will perform a one-time preprocessing using $O(MN)$ time. After this preprocessing, for every slid-pair candidate, we can extract m right-matching messages for f in $O(m) = O(M^{2/3}N^{1/6})$ time, and even better, those messages are already grouped according to their right segments. The same running time would be needed to extract messages for g. We then can run LHD to eliminate most false slid-pair candidates.

Proceeding into details, suppose that we have a $\mathtt{U}$-chain $U_0 \to U_1 \to \cdots \to U_{2p}$ and a $\mathtt{V}$-chain $V_0 \to V_1 \to \cdots \to V_{2p}$, and we want to check if (U_i, V_0) is a slid pair, for every $k \leq \{0, 1, \ldots, p-1\}$. For each slid-pair candidate (U_i, V_0), the known plaintext/ciphertext pairs for f are (U_{i+k}, V_k) for $k \leq 2p - i$, and the known plaintext/ciphertext pairs for g are (V_k, U_{i+k+1}), for $k \leq 2p - k - 1$.

- In order to preprocess these p slid-pair candidates for f, note that they all use plaintexts $U_{p+1}, \ldots, U_{2p}$. So we will partition these plaintexts by their right

Procedure $\mathsf{Slide}(\boldsymbol{U}, \boldsymbol{V})$
$(U_0, \ldots, U_{2p}) \leftarrow \boldsymbol{U}$; $(V_0, \ldots, V_{2p}) \leftarrow \boldsymbol{V}$; $(Z_1, \ldots, Z_{MN}) \leftarrow \mathbb{Z}_M \times \mathbb{Z}_N$
$L \leftarrow \mathsf{Process}((U_{p+1}, p+1), \ldots, (U_{2p}, 2p))$
$L' \leftarrow \mathsf{Process}((V_0, 0), \ldots, (V_{p-1}, p-1))$
For $i = 0$ to $p-1$ do // Check if (U_i, V_0) is a slid pair
  If $(\mathsf{Dist}(L, \boldsymbol{V}, -i) \wedge \mathsf{Dist}(L', \boldsymbol{U}, i+1))$ then
    $\boldsymbol{X} \leftarrow (U_p, \ldots, U_{2p-1})$; $\boldsymbol{Y} \leftarrow (V_{p-i}, \ldots, V_{2p-1-i})$
    $\boldsymbol{X}^* \leftarrow (V_0, \ldots, V_{p-1})$; $\boldsymbol{Y}^* \leftarrow (U_{i+1}, \ldots, U_{i+p})$
    $f \leftarrow\!\!\$\ \mathsf{Recover}(\boldsymbol{X}, \boldsymbol{Y})$; $g \leftarrow\!\!\$\ \mathsf{Recover}(\boldsymbol{X}^*, \boldsymbol{Y}^*)$
    If $f \neq \perp$ and $g \neq \perp$ then
      For $i = 1$ to MN do $C_i \leftarrow g(f(Z_i))$
      Return $(C_1, \ldots, C_{MN})$ // Codebook is $(Z_1, C_1), \ldots, (Z_{MN}, C_{MN})$

Procedure $\mathsf{Process}((X_1, r_1), \ldots, (X_p, r_p))$ // Preprocessing
$L \leftarrow \emptyset$; $m \leftarrow \lceil \frac{p}{N} \cdot \lceil 32N^{1/6} \rceil \rceil$
Partition $(X_1, r_1), \ldots, (X_p, r_p)$ by the right segments
Let $P_1, \ldots, P_M$ be the resulting partitions, with $|P_1| \geq \cdots \geq |P_M|$
For $i = 1$ to M, $(X, r) \in P_i$ do: If $|L| < m$ then $L \leftarrow L \cup \{(X, r)\}$
Return L

Procedure $\mathsf{Dist}(L, \boldsymbol{V}, k)$ // Running LHD with preprocessed list L
$i \leftarrow 1$, $(V_0, \ldots, V_{2p}) \leftarrow \boldsymbol{V}$; $m \leftarrow |L|$
For $(Z, r) \in L$ do $X_i \leftarrow Z$; $C_i \leftarrow V_{r+k}$; $i \leftarrow i + 1$
Return $\mathsf{LHD}(X_1, \ldots, X_m, C_1, \ldots, C_m)$

Procedure $\mathsf{Recover}(\boldsymbol{X}, \boldsymbol{Y})$
$(X_0, \ldots, X_{p-1}) \leftarrow \boldsymbol{X}$; $(Y_0, \ldots, Y_{p-1}) \leftarrow \boldsymbol{Y}$
$(F_1, F_2, F_3, F_4) \leftarrow\!\!\$\ \mathsf{TF}(X_0, \ldots, X_{p-1}, Y_0, \ldots, Y_{p-1})$
Let f be the 4-found Feistel of round functions (F_1, F_2, F_3, F_4)
// Function f will be $\perp$ if TF does not fully recover (F_1, F_2, F_3, F_4)
Return f

Fig. 7. The implementation of procedure Slide in the SD attack. The numbers M and N are global parameters. We assume that there is a global total ordering on the domain $\mathbb{Z}_M \times \mathbb{Z}_N$, so that we can write $(Z_1, \ldots, Z_{MN}) \leftarrow \mathbb{Z}_M \times \mathbb{Z}_N$.

segments into (possibly empty) groups $P_1, \ldots, P_M$, with $|P_1| \geq \cdots \geq |P_M|$. We then store m messages U_j from the first $\lceil 32N^{1/6} \rceil$ partitions, together with their indices j, in a list L. Later, for a slid-pair candidate (U_i, V_0), we iterate through pairs (U_j, j) in L, and for each such pair, the corresponding ciphertext of U_j for f is V_{j-i}, which takes $O(1)$ time to find if we store $(U_0, \ldots, U_{2p})$ and $(V_0, \ldots, V_{2p})$ in arrays.

– Preprocessing for g is similar, but note that the p candidates all use plaintexts $V_0, V_1, \ldots, V_{p-1}$.

For a pair of U chain and V chain, partitioning takes $O(p + M)$ time, and by using a max-heap, we can find the $\lceil 32N^{1/6} \rceil$ biggest partitions in $O(M)$ time, and extracting m messages from those partitions takes $O(m)$ time. Summing up, for a pair of U chain and V chain, the running time of the preprocessing is $O(p)$. Hence, totally, for s^2 pairs of U chains and V chains, the overall running time of the preprocessing is $O(s^2 p) = O(MN)$.

Putting things together. By combining the LHD attack and the preprocessing, one can implement procedure Slide as in Fig. 7. Thus the SD attack uses $O(sp) = O(M^{5/6}N)$ queries and space, and its running time is $O(MN)$ for the preprocessing, $O(MN \cdot m) = O(M^{5/3}N^{7/6})$ for running LHD, and expectedly, $O(M^{5/3}N)$ for running TF. Hence the total running time of SD is $O(M^{5/3}N^{7/6})$.

Improving the recovery rate. To improve the recovery rate of SD, one can run the attack several times with different tweak pairs $(T_1, T_1 \oplus \mathsf{Mask}), \ldots, (T_r, T_r \oplus \mathsf{Mask})$, where $\mathsf{Mask} = [4]_\tau \,\|\, [4]_\tau$. If $(T_i \oplus T_j)[1\colon \tau], (T_i \oplus T_j)[\tau + 1\colon 2\tau] \notin \{[0]_\tau, \ldots, [7]_\tau\}$ for every $i \neq j$ then those r instances of SD will call AES on different τ-bit prefixes. If we model AES as a good PRF then the results of those SD instances are independent. Hence if the recovery rate of SD is ϵ then running it for r times will have recovery rate $1 - (1 - \epsilon)^r$.

4 Attacking 4-round Feistel-based Blockciphers

In this section, we generalize and improve DV's known-plaintext codebook-recovery attack on Feistel-based, 4-round blockciphers where the Feistel network might be unbalanced. In particular, on a four-round balanced Feistel of domain size N^2, DV's attack runs in $O(N^3)$ expected time, but our attack, which we name *Triangle-Finding* (TF), runs in only $O(N^{5/3})$ expected time. Both our attack and DV's rely on a conjecture of Feistel networks that DV empirically verified for balanced Feistel of domain $\{0,1\}^{2n}$, for $n \in \{1, \ldots, 9\}$. We prove that this conjecture indeed holds, making both attacks unconditional.

Let

$$p = \max\left\{ \lfloor 2^{1/3} M^{2/3} N \rfloor, \lceil M(\ln(M) + 5) \rceil \right\} .$$

In our attack, we suppose that we are given p pairs $(X_1, C_1), \ldots, (X_p, C_p)$ of plaintext/ciphertext under a four-round Feistel network $\mathsf{F} = \mathbf{Feistel}[4, M, N, \boxplus]$ with a uniformly random key, where $M \geq N \geq 64$ and the plaintexts are chosen uniformly without replacement from $\mathbb{Z}_M \times \mathbb{Z}_N$. To simplify our asymptotic analysis, we assume that $N = \Omega(\sqrt{M})$, which applies to the setting of the FF3 scheme, since FF3 uses near-balanced Feistel. Thus $p \in O(M^{2/3}N)$.

In our attack, we need a graph representation of the first p plaintext/ciphertext pairs, and a few algorithms, which we will elaborate in Section 4.1. We then describe our TF attack in Section 4.2.

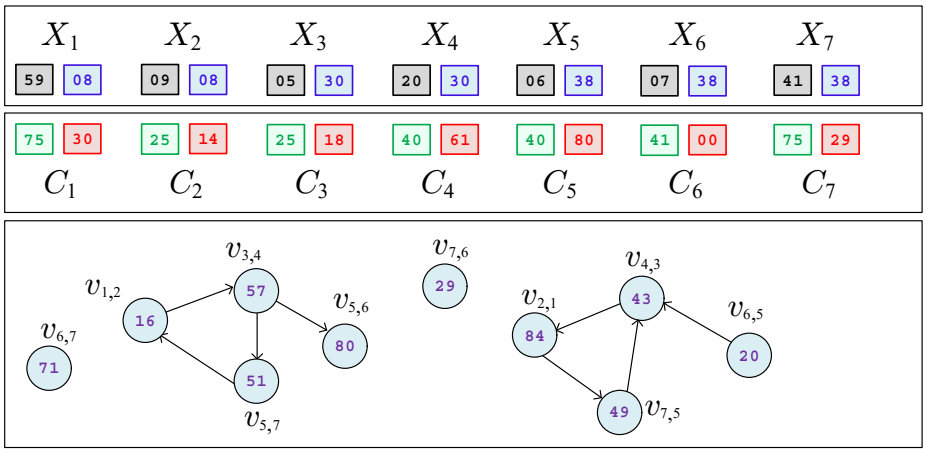

Fig. 8. Top: Seven pairs of plaintext/ciphertext $(X_1, C_1), \ldots, (X_7, C_7)$ on $\mathbb{Z}_M \times \mathbb{Z}_N$, with $M = N = 100$. **Bottom:** Differential graph $\mathcal{G}$ constructed from those seven pairs.

4.1 Differential Graph and Its Triangles

Differential graph. Let $\mathcal{G} = (V, E)$ be the following directed graph. First, for each $i \neq j$, we create a node $v_{i,j}$ with label $\mathsf{Label}(v_{i,j}) = \mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_j)$ if $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$ and $\mathsf{LH}(C_i) \boxminus \mathsf{LH}(X_i) = \mathsf{LH}(C_j) \boxminus \mathsf{LH}(X_j)$. Next, for every two nodes $v_{i,j}$ and $v_{k,\ell}$ such that i, j, k, ℓ are distinct, we create a directed edge $(v_{i,j}, v_{k,\ell})$ if $\mathsf{LH}(C_j) = \mathsf{LH}(C_k)$ and the following non-degeneracy conditions hold:

(1) $\mathsf{LH}(C_i) \neq \mathsf{LH}(C_\ell)$,
(2) $\mathsf{RH}(X_i) \neq \mathsf{RH}(X_k)$,
(3) $\mathsf{RH}(X_j) \boxminus \mathsf{RH}(X_k) \neq \mathsf{RH}(C_j) \boxminus \mathsf{RH}(C_k)$, and
(4) $\mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j) \neq \mathsf{LH}(X_k) \boxminus \mathsf{LH}(X_\ell)$.

See Fig. 8 for an illustration of the graph $\mathcal{G}$. We call $\mathcal{G}$ the *differential graph* of the plaintext/ciphertexts pairs $(X_1, C_1), \ldots, (X_p, C_p)$.

Collision. For each plaintext X_i, let X_i^t denote the round-t intermediate output of X_i under F. We say that X_i and X_j *collide* at round t, if either (i) t is odd and $\mathsf{LH}(X_i^t) = \mathsf{LH}(X_j^t)$, or (ii) t is even and $\mathsf{RH}(X_i^t) = \mathsf{RH}(X_j^t)$. Lemma 7 below, by Hoang and Rogaway [19], shows that for two distinct, non-adaptive queries, collision at the first round is unlikely.

Lemma 7. [19] *Let $M, N, r \geq 1$ be integers. Let $\mathsf{F} = \mathbf{Feistel}[r, M, N, \boxplus]$. For any distinct, non-adaptive messages X and X', the chance that they collide at round 1 is at most $1/M$. Moreover, if X and X' have the same right segment then they will certainly not collide at round 1.*

Lemma 8 below characterizes collisions in the differential graph. The proof is in Appendix C.7.

Lemma 8. *Let $M, N \geq 10$ be integers and let $\mathsf{F} = \mathbf{Feistel}[4, M, N, \boxplus]$. Let $(X_1, C_1), \ldots, (X_p, C_p)$ be plaintext/ciphertext pairs under F with a uniformly random key, where the plaintexts are chosen uniformly without replacement from $\mathbb{Z}_M \times \mathbb{Z}_N$. Let $\mathcal{G} = (V, E)$ be the differential graph of those pairs, and let $(v_{i,j}, v_{k,\ell})$ be an edge of $\mathcal{G}$. Partition the four corresponding messages into two groups: $\{X_i, X_j\}, \{X_k, X_\ell\}$ Then*

(a) There is no intra-group collision at round 1.
(b) X_k and X_j do no collide at round 1.
(c) There is at most one inter-group collision at round 1.
(d) There is no inter-group collision at round 2.

GOOD VERSUS BAD NODES. For a node $v_{i,j}$ in the differential graph $\mathcal{G}$, we say that it is *good* if $\mathsf{RH}(X_i^2) = \mathsf{RH}(X_j^2)$; otherwise we say that it is *bad*. Lemma 9 below characterizes an important property of good nodes; it is a direct generalization of a result in DV's work.

Lemma 9. *Let $M, N \geq 10$ be integers and let $\mathsf{F} = \mathbf{Feistel}[4, M, N, \boxplus]$. Let $(X_1, C_1), \ldots, (X_p, C_p)$ be plaintext/ciphertext pairs under F with a uniformly random key, where the plaintexts are chosen uniformly without replacement from $\mathbb{Z}_M \times \mathbb{Z}_N$. Let $\mathcal{G} = (V, E)$ be the differential graph of those pairs, and let (G_1, G_2, G_3, G_4) be the functions specified by the secret key of F. Then for any good node $v_{i,j} \in V$, we have $\mathsf{Label}(v_{i,j}) = G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_j))$.*

Proof. Recall that $\mathsf{Label}(v_{i,j}) = \mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_j)$. Next, since

$$\mathsf{RH}(C_i) = G_4(\mathsf{LH}(X_i^3)) \boxplus \mathsf{RH}(X_i^3)$$

and since $\mathsf{LH}(X_i^3) = \mathsf{LH}(C_i)$ and $\mathsf{RH}(X_i^3) = \mathsf{RH}(X_i^2)$,

$$\mathsf{RH}(C_i) = G_4(\mathsf{LH}(C_i)) \boxplus \mathsf{RH}(X_i^2) \ . \tag{2}$$

Likewise,

$$\mathsf{RH}(C_j) = G_4(\mathsf{LH}(C_j)) \boxplus \mathsf{RH}(X_j^2) \ . \tag{3}$$

Subtracting Equation (2) and Equation (3) side by side, and note that $\mathsf{RH}(X_i^2) = \mathsf{RH}(X_j^2)$ since $v_{i,j}$ is good, we obtain

$$\mathsf{Label}(v_{i,j}) = G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_j))$$

as claimed. □

The following Lemma 10 computes the average number of good and bad nodes, and estimates the average number of edges in the differential graph; see Appendix C.8 for the proof.

Lemma 10. *Let $M, N \geq 10$ be integers and let $\mathsf{F} = \mathbf{Feistel}[4, M, N, \boxplus]$. Let $(X_1, C_1), \ldots, (X_p, C_p)$ be plaintext/ciphertext pairs under F with a uniformly random key, where the plaintexts are chosen uniformly without replacement from $\mathbb{Z}_M \times \mathbb{Z}_N$. Let $\mathcal{G} = (V, E)$ be the differential graph of those pairs, and let Z be the random variable for the number of good nodes in $\mathcal{G}$. Then*

(a) $\mathbf{E}[|V|] = \frac{p(p-1)(M-1)(M+N-1)}{MN(MN-1)}$.
(b) $\mathbf{E}[Z] = \frac{p(p-1)(M-1)}{(MN-1)N}$.
(c) $\mathbf{E}[|E|] \le \frac{p!}{(p-4)!} \cdot \frac{(M+N)^2}{M^3N^4}$.

Since $p = O(M^{2/3}N)$ and $M \ge N$, from Lemma 10, on average, the differential graph $\mathcal{G}$ contains about $O(M^{4/3})$ nodes, and the majority of them are good. In addition, there are on average $O(M^{5/3})$ edges in $\mathcal{G}$.

A FAST CONSTRUCTION OF DIFFERENTIAL GRAPHS. The naive approach to construct $\mathcal{G}$ would take $\Theta(p^2) = \Theta\big(M^{10/3}\big)$ time just to construct the node set V. We now show how to build $\mathcal{G}$ in $O\big(M^{5/3}\big)$ expected time; the code is given in Fig. 9.

- First, partition the pairs (X_i, C_i) based on $(\mathsf{LH}(C_i) \boxminus \mathsf{LH}(X_i), \mathsf{RH}(X_i))$. Using appropriate data structure, this takes $O(p)$ time.
- For each partition P, enumerate all distinct pairs $(X_i, C_i), (X_j, C_j) \in P$. Each such pair forms a node in V, and its label can be computed accordingly. This takes $O(|V|)$ time.
- Finally, partition V into M groups P_d with $d \in \mathbb{Z}_M$, such that each node $v_{i,j}$ goes to group $P_{\mathsf{LH}(C_j)}$. Also, partition V into M groups S_d with $d \in \mathbb{Z}_M$, such that each node $v_{k,\ell}$ goes to group $S_{\mathsf{LH}(C_k)}$. By enumerating elements in $P_d \times S_d$ (with some pruning) for every $d \in \mathbb{Z}_M$ via appropriate data structure, we can create the edge set E using

$$O\Big(|V| + M + \sum_{i,j,k,\ell} D_{i,j,k,\ell}\Big)$$

  time, where $D_{i,j,k,\ell}$ is the indicator random variable for the event that (i) $v_{i,j} \in V$, (ii) $v_{k,\ell} \in V$, and (iii) $\mathsf{LH}(C_j) = \mathsf{LH}(C_k)$. The summation is taken over all distinct $i, j, k \in \{1, \ldots, p\}$ and $\ell \in \{1, \ldots, p\} \backslash \{k\}$. By pretending that (i), (ii), (iii) are independent, we can heuristically estimate that $\Pr[D_{i,j,k,\ell} = 1] \lessapprox \frac{4}{MN^4}$.

Hence the total expected time is

$$O\Big(p + M + \mathbf{E}[|V|] + \frac{4p^4}{MN^4}\Big) = O\big(M^{5/3}\big) \ .$$

TRIANGLES. Recall that from Lemma 10, on average, the majority of nodes in the differential graph are good. We now describe a method to realize good nodes with high probability. A *triangle* of the graph $\mathcal{G}$ is a directed cycle of length 3. For a triangle $\mathcal{T} = (u_1, u_2, u_3, u_1)$, its *weight* $\mathsf{weight}(\mathcal{T})$ is defined as the sum of the labels, meaning that

$$\mathsf{weight}(\mathcal{T}) = \mathsf{Label}(u_1) \boxplus \mathsf{Label}(u_2) \boxplus \mathsf{Label}(u_3) \ .$$

DV observed that in the balanced setting, for a triangle $\mathcal{T}$, if all of its three nodes are good then its weight is 0. Lemma 11 below shows that their observation also holds in the unbalanced setting.

Procedure $\mathsf{BuildGraph}(X_1, C_1, \ldots, X_p, C_p)$

$V, E \leftarrow \emptyset$; Initialize a hash table H
For $i = 1$ to p do $Z \leftarrow (\mathsf{LH}(C_i) \boxminus \mathsf{LH}(X_i), \mathsf{RH}(X_i))$; $P \leftarrow H[Z]$; $P \leftarrow P \cup \{i\}$
For each key Z in H do
  For $i, j \in H[Z]$, with $i \neq j$, do
    $v_{i,j} \leftarrow (i, j)$; $\mathsf{Label}(v_{i,j}) \leftarrow \mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_j)$; $V \leftarrow V \cup \{v_{i,j}\}$
For $d \in \mathbb{Z}_M$ do $P_d, S_d \leftarrow \emptyset$
For $v \in V$ do $(i, j) \leftarrow v$; $P_{\mathsf{LH}(C_j)} \leftarrow P_{\mathsf{LH}(C_j)} \cup \{v\}$; $P_{\mathsf{LH}(C_i)} \leftarrow P_{\mathsf{LH}(C_i)} \cup \{v\}$
For $d \in \mathbb{Z}_M$ do
  Initialize a hash table H_d
  For $v = (k, \ell) \in S_d$ do $L \leftarrow H_d[k]$; $L \leftarrow L \cup \{v\}$
  For every $u = (i, j) \in P_d$ and every key $k \in H_d$ such that $k \notin \{i, j\}$ do
    For every $v = (k, \ell) \in H_d[k]$ such that $\ell \notin \{i, j\}$ do
      // Check non-degeneracy requirements
      If $\mathsf{LH}(C_i) \neq \mathsf{LH}(C_\ell)$ and $\mathsf{RH}(X_i) \neq \mathsf{RH}(X_k)$
        and $\mathsf{RH}(X_j) \boxminus \mathsf{RH}(X_k) \neq \mathsf{RH}(C_j) \boxminus \mathsf{RH}(C_k)$
        and $\mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j) \neq \mathsf{LH}(X_k) \boxminus \mathsf{LH}(X_\ell)$ then $E \leftarrow E \cup \{(u, v)\}$
Return $\mathcal{G} = (V, E)$

Fig. 9. Code for building the differential graph $\mathcal{G} = (V, E)$ of $X_1, C_1, \ldots, X_p, C_p$.

Lemma 11. *Let $M, N \geq 10$ be integers and let $\mathsf{F} = \mathbf{Feistel}[4, M, N, \boxplus]$. Let $(X_1, C_1), \ldots, (X_p, C_p)$ be plaintext/ciphertext pairs under F with a uniformly random key, where the plaintexts are chosen uniformly without replacement from $\mathbb{Z}_M \times \mathbb{Z}_N$. Let $\mathcal{G} = (V, E)$ be the differential graph of those pairs. For a triangle $\mathcal{T}$ in $\mathcal{G}$, if all the three nodes of $\mathcal{T}$ are good then $\mathsf{weight}(\mathcal{T}) = 0$.*

Proof. Consider a triangle $(v_{i,j}, v_{k,\ell}, v_{r,s})$ such that all the three nodes are good. Let (G_1, G_2, G_3, G_4) be the functions specified by the secret key of F. From Lemma 9,

$$\mathsf{Label}(v_{i,j}) = G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_j)) \ ,$$
$$\mathsf{Label}(v_{k,\ell}) = G_4(\mathsf{LH}(C_k)) \boxminus G_4(\mathsf{LH}(C_\ell)) \ ,$$
$$\mathsf{Label}(v_{r,s}) = G_4(\mathsf{LH}(C_r)) \boxminus G_4(\mathsf{LH}(C_s)) \ .$$

Since $(v_{i,j}, v_{k,\ell}, v_{r,s})$ is a directed cycle in $\mathcal{G}$, $\mathsf{LH}(C_k) = \mathsf{LH}(C_j)$ and $\mathsf{LH}(C_r) = \mathsf{LH}(C_\ell)$ and $\mathsf{LH}(C_s) = \mathsf{LH}(C_i)$. Thus the sum of the three labels is indeed 0. □

Above, we show that a triangle whose nodes are all good will have weight 0. The following Lemma 12 shows that the converse holds with very high probability; see Appendix C.9 for the proof. This proves a conjecture in DV's work [15] that they empirically verified for the balanced, boolean case $M = N = 2^n$, with $n \in \{2, 3, \ldots, 9\}$. We also give a rigorous lower bound for the expected number

of triangles whose all nodes are good, whereas DV could only give a heuristic estimation of this number.

Lemma 12. *Let $M, N \geq 19$ be integers and let $\mathsf{F} = \mathbf{Feistel}[4, M, N, \boxplus]$. Let $(X_1, C_1), \ldots, (X_p, C_p)$ be plaintext/ciphertext pairs under F with a uniformly random key, where the plaintexts are chosen uniformly without replacement from $\mathbb{Z}_M \times \mathbb{Z}_N$. Let $\mathcal{G} = (V, E)$ be the differential graph of those pairs.*

(a) For a triangle $\mathcal{T}$ in $\mathcal{G}$ of zero weight, the probability that all nodes in $\mathcal{T}$ are good is at least $\frac{1}{1+\epsilon}$, where $\epsilon = \frac{N}{M-9}\left(\frac{4}{M} + \frac{33N}{(N-2)M^2} + \frac{39}{M^2}\right)$.
(b) The expected number of triangles in $\mathcal{G}$ whose all nodes are good is at least $\frac{p!}{3(p-6)!} \cdot \frac{1}{M^3N^6}\left(1 - \frac{7}{N} - \frac{14}{M}\right)$.

Discussion. While the core idea of differential graphs is from DV's work, there are important differences between our definition and DV's:

- Because of the symmetry of Feistel, one can view $\mathsf{Rev}(X_1), \ldots, \mathsf{Rev}(X_p)$ as the "ciphertexts" of $\mathsf{Rev}(C_1), , \ldots, \mathsf{Rev}(C_p)$ under a four-round Feistel $\overline{\mathsf{F}} = \mathbf{Feistel}[4, N, M, \boxminus]$ where $\mathsf{Rev}(X) = (R, L)$ for any $X = (L, R) \in \mathbb{Z}_M \times \mathbb{Z}_N$. In this sense, DV's notion is the dual of ours. While the two definitions yield no difference in the balanced setting, if $M \gg N$, DV's notion would give a much poorer bound[10] in a dual version of Lemma 12, leading to an inferior recovery rate of the TF attack.
- Our notion also adds some non-degeneracy requirements, allowing us to prove DV's conjecture on differential graph in Lemma 12 further below.

Enumerating zero-weight triangles. From Lemma 12, a simple way to realize good nodes is to enumerate all triangles of zero weight. We now show how to do that in $O(M^{5/3})$ expected time; the code is given in Fig. 10.

- First, for each node $v \in V$, partition its set of incoming edges such that each edge (u, v) goes to group $P_{v,d}$, with $d = \mathsf{Label}(u) \boxplus \mathsf{Label}(v)$, and also partition the set of outgoing edges into groups such that each edge (v, w) goes to group $S_{v,s}$, where $s = 0 \boxminus \mathsf{Label}(w)$.
- Next for each (v, d), enumerate all pairs $(u, w) \in P_{v,d} \times S_{v,d}$ such that there is a directed edge $(w, u) \in E$. Each triple (v, u, w) is a triangle of zero weight.

Using appropriate data structure, the first step takes $O(|V| + |E|)$ time, whereas the cost of the second step is in the order of

$$\sum_{v \in V} \sum_{d \in \mathbb{Z}_N} |P_{v,d}| \cdot (1 + |S_{v,d}|) \leq \left(\sum_{v \in V} \sum_{d \in \mathbb{Z}_N} |P_{v,d}|\right) + \left(\sum_{v \in V} \sum_{d \in \mathbb{Z}_N} |P_{v,d}| \cdot |S_{v,d}|\right)$$

[10] In fact, the dual version of Lemma 12 would yield the bound $\frac{1}{1+\epsilon^*}$, where $\epsilon^* = \frac{M}{N-9}\left(\frac{4}{N} + \frac{33M}{(M-2)N^2} + \frac{39}{N^2}\right)$. Concretely, for $M = 100$ and $N = 10$, the bound in our Lemma 12 is 0.9947, whereas its dual is much poorer, just 0.0089.

Procedure $\mathsf{GetTriangles}(\mathcal{G}, X_1, C_1, \ldots, X_p, C_p)$
$(V, E) \leftarrow \mathcal{G}$; $L \leftarrow \emptyset$; Initialize hash tables H_0, H_1, H_2
For every edge $(u, v) \in E$ do
  $d \leftarrow \mathsf{Label}(u) \boxplus \mathsf{Label}(v)$; $P \leftarrow H_1[v, d]$; $P \leftarrow P \cup \{e\}$
  $s \leftarrow 0 \boxminus \mathsf{Label}(v)$; $S \leftarrow H_2[u, s]$; $S \leftarrow S \cup \{e\}$; $H_0[e] \leftarrow 1$
For every key (v, d) of H_1, every $u \in H_1[v, d], w \in H_2[v, d]$ do
  If $H_0[(w, u)] = 1$ then $\mathcal{T} \leftarrow (u, v, w)$; $L \leftarrow L \cup \{\mathcal{T}\}$
Return L

Fig. 10. Code to enumerate triangles of zero weight from the differential graph $\mathcal{G} = (V, E)$ of $X_1, C_1, \ldots, X_p, C_p$.

$$= \Big(\sum_{v \in V} \mathsf{indeg}(v)\Big) + \Big(\sum_{v \in V} \sum_{d \in \mathbb{Z}_N} |P_{v,d}| \cdot |S_{v,d}|\Big)$$
$$= |E| + \Big(\sum_{v \in V} \sum_{d \in \mathbb{Z}_N} |P_{v,d}| \cdot |S_{v,d}|\Big) \ ,$$

where $\mathsf{indeg}(v)$ is the incoming degree of node v. Since $\mathbf{E}[|V|] \in O(M^{4/3})$ and $\mathbf{E}[|E|] \in O(M^{5/3})$, what remains is to show that

$$\mathbf{E}\Big[\sum_{v \in V} \sum_{d \in \mathbb{Z}_N} |P_{v,d}| \cdot |S_{v,d}|\Big] \in O(M^{5/3}) \ . \tag{4}$$

For each tuple $\mathcal{L} = (i, j, k, \ell, r, s, d) \in (\{1, \ldots, p\})^6 \times \mathbb{Z}_N$ such that i, j, k, ℓ, r, s are distinct, let $B_{\mathcal{L}}$ denote the Bernoulli random variable such that $B_{\mathcal{L}} = 1$ if and only if $(v_{i,j}, v_{k,\ell})$ and $(v_{r,s}, v_{i,j})$ are edges of $\mathcal{G}$, and $v_{k,\ell} \in S_{v_{i,j},d}$ and $v_{r,s} \in P_{v_{i,j},d}$. Then

$$\mathbf{E}\Big[\sum_{v \in V} \sum_{d \in \mathbb{Z}_N} |P_{v,d}| \cdot |S_{v,d}|\Big] = \mathbf{E}\Big[\sum_{\mathcal{L}} B_{\mathcal{L}}\Big] = \sum_{\mathcal{L}} \mathbf{E}[B_{\mathcal{L}}] = \sum_{\mathcal{L}} \Pr[B_{\mathcal{L}} = 1] \ .$$

Note that for each $\mathcal{L} = (i, j, k, \ell, r, s, d)$, the event $B_{\mathcal{L}} = 1$ happens only if the following events happen: (1) $v_{i,j} \in V$, (2) $v_{j,k} \in V$, (3) $v_{r,s} \in V$, (4) $\mathsf{LH}(C_j) = \mathsf{LH}(C_k)$, (5) $\mathsf{LH}(C_s) = \mathsf{LH}(C_i)$, (6) $\mathsf{Label}(v_{i,j}) \boxplus \mathsf{Label}(v_{r,s}) = d$, and (7) $\mathsf{Label}(v_{k,\ell}) = 0 \boxminus d$. By pretending that these seven events are independent, from Lemma 3, we can heuristically estimate

$$\Pr[B_{\mathcal{L}} = 1] \lessapprox \Big(\frac{M + N - 1}{MN} \cdot \frac{1}{N}\Big)^3 \Big(\frac{1}{M}\Big)^2 \Big(\frac{1}{N}\Big)^2 \leq \frac{(M + N)^3}{M^5 N^8} \leq \frac{8}{M^2 N^8} \ .$$

Hence

$$\sum_{\mathcal{L}} \mathbf{E}[B_{\mathcal{L}}] \lessapprox \frac{p! \cdot N}{(p - 6)!} \cdot \frac{8}{M^2 N^8} \in O(M^2/N) \ .$$

Moreover, due to our assumption that $N = \Omega(\sqrt{M})$, it follows that $M^2/N \in O(M^{1.5})$. We then conclude that

$$\mathbf{E}\Big[\sum_{v \in V} \sum_{d \in \mathbb{Z}_N} |P_{v,d}| \cdot |S_{v,d}|\Big] \in O(M^{1.5})$$

and thus justify Equation (4).

REMARKS. DV also considered the problem of finding zero-weight triangles for the balanced case $M = N$. They first enumerated all triangles via sparse matrix multiplications, and then computed the sum of labels for each triangle. In this balanced setting, DV's algorithm takes $O(N^3)$ time, whereas our algorithm takes $O(N^{5/3})$ time.

4.2 The TF Attack

We begin with a simple but useful observation of DV on four-round Feistel.

AN OBSERVATION. Let (F_1, F_2, F_3, F_4) be the round functions of F. For any $\Delta \in \mathbb{Z}_N$, let $\mathsf{Shift}(\mathsf{F}, \Delta)$ denote a 4-round Feistel network $\overline{\mathsf{F}} = \mathbf{Feistel}[4, M, N, \boxplus]$ of round functions $(\overline{F}_1, \overline{F}_2, \overline{F}_3, \overline{F}_4)$ such that $\overline{F}_1 = F_1$, $\overline{F}_2(K, x) = F_2(K, x) \boxminus \Delta$, $\overline{F}_3(K, y) = F_3(K, y \boxplus \Delta)$, and $\overline{F}_4(K, x) = F_4(K, x) \boxplus \Delta$, for any $x \in \mathbb{Z}_M$, $y \in \mathbb{Z}_N$, and any key K. Note that for any choice of Δ, scheme $\overline{\mathsf{F}} = \mathsf{Shift}(\mathsf{F}, \Delta)$ ensures that $\overline{\mathsf{F}}.\mathsf{E}(K, X) = \mathsf{F}.\mathsf{E}(K, X)$ for any key K and any $X \in \mathbb{Z}_M \times \mathbb{Z}_N$. Therefore, in a codebook recovery attack against $\mathsf{F}.\mathsf{E}(K, \cdot)$, without loss of generality, one can pick a designated point $x^* \in \mathbb{Z}_M$ and *assume* that $F_4(K, x^*) = 0$.

THE ATTACK. Let (G_1, G_2, G_3, G_4) be the functions specified by the secret key of F. We will recover even the tables of those functions, instead of just the codebook.

Our attack TF is based on a known-plaintext codebook-recovery attack RY on three-round Feistel that we describe in Appendix B. If we run RY on $\ell \geq \max\{\lceil N(\ln(N)+\ln(2)+\lambda)\rceil, \lceil M(\ln(M)+\delta)\rceil\}$ known plaintext/ciphertext pairs, for any $\lambda, \delta > 0$, the RY attack will take $O(\ell)$ time, and recovers all the round functions of the three-round Feistel with probability around $e^{-e^{-\lambda}} - e^{-\delta}$. If we just have $\ell \geq \lceil N(\ln(N)+\ln(2)+\lambda)\rceil$ then RY will recover the top round function with probability at least $e^{-e^{-\lambda}}$. We note that RY is used in a modular way; one does not need to know the technical details of RY to understand the TF attack.

While TF somewhat resembles DV's attack on 4-round Feistel, there are important changes to improve efficiency and recovery rate, which we will elaborate further below. The code of TF is given in Fig. 11; below we will describe the attack.

In the TF attack, we will first construct the differential graph $\mathcal{G} = (V, E)$ of the plaintext/ciphertext pairs $(X_1, C_1), \ldots, (X_p, C_p)$, and then enumerate all triangles of $\mathcal{G}$ of zero weight. Let S be the set of the nodes of those triangles; each node S is very likely to be good, due to Lemma 12. For each $v_{i,j} \in S$, from Lemma 9, if $v_{i,j}$ is indeed good then $\mathsf{Label}(v_{i,j}) = G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_j))$.

Our first step is to recover several (but possibly not all) entries of G_4. In order to do that, construct the following undirected graph $\mathcal{G}^* = (V^*, E^*)$ of $|V^*| = M$ nodes. Nodes in V^* are distinctly labeled by elements of $\mathbb{Z}_M$. For each node $v_{i,j} \in S$, we create an edge between nodes $\mathsf{LH}(C_i)$ and $\mathsf{LH}(C_j)$ of V^*, indicating that we know the difference between $G_4(\mathsf{LH}(C_i))$ and $G_4(\mathsf{LH}(C_j))$. Once the

Procedure $\mathsf{TF}(X_1, C_1, \ldots, X_p, C_p)$

$\mathcal{G} \leftarrow \mathsf{BuildGraph}(X_1, C_1, \ldots, X_p, C_p)$ // Build the differential graph
$L \leftarrow \mathsf{GetTriangles}(\mathcal{G}, X_1, C_1, \ldots, X_p, C_p)$ // Enumerate zero-weight triangles
$V^*, E^*, S \leftarrow \emptyset$; Initialize a hash table H
For every $\mathcal{T} \in L$ do $(u, v, w) \leftarrow \mathcal{T}$; $S \leftarrow S \cup \{u, v, w\}$
For $i \in \mathbb{Z}_M$ do $V^* \leftarrow V^* \cup \{i\}$
For every $v \in S$ do
  $(i, j) \leftarrow v$; $e \leftarrow \{\mathsf{LH}(C_i), \mathsf{LH}(C_j)\}$; $E^* \leftarrow E^* \cup \{e\}$
  $H[e] \leftarrow (\mathsf{LH}(C_i), \mathsf{Label}(v))$
$\mathcal{G}^* \leftarrow (V^*, E^*)$; Let $\mathcal{C}$ be the biggest connected component of $\mathcal{G}^*$
For $i \leftarrow 1$ to μ do
  $(G_1, G_2, G_3, G_4) \leftarrow \mathsf{Restore}(X_1, C_1, \ldots, X_p, C_p, \mathcal{C})$
  For $j \leftarrow 1$ to $3M$ do // Checking consistency of (G_1, G_2, G_3, G_4)
    $(L, R) \leftarrow X_j$
    For $k \leftarrow 1$ to 4 do
      If $(k \bmod 2 = 1)$ then $L \leftarrow L \boxplus G_k(R)$ else $R \leftarrow R \boxplus G_k(L)$
    $C_j^* \leftarrow (L, R)$
  If $(C_1, \ldots, C_{3M}) = (C_1^*, \ldots, C_{3M}^*)$ then return (G_1, G_2, G_3, G_4)

Fig. 11. The TF attack (parameterized by a small number μ) on 4-round Feistel, which is based on another attack RY on 3-round Feistel and a procedure Restore in Fig. 12.

graph $\mathcal{G}^*$ is constructed, we pick an arbitrary node $x^* \in V^*$ that belongs to the biggest connected component of $\mathcal{G}^*$, and set $G_4(x^*) \leftarrow 0$. We then recover $G_4(u)$ for every node $u \in V^*$ reachable from x^* using breadth-first search (BFS), but stop when $\left\lfloor \frac{3M}{\sqrt{N}} \right\rfloor$ entries of G_4 are recovered. Let $\mathcal{I} \subseteq \{1, 2, \ldots, p\}$ be the set of indices i such that $G_4(\mathsf{LH}(C_i))$ is recovered at this point.

Our next step is to recover the entire table of G_1 using RY. For each $i \in \mathcal{I}$, recover the round-3 intermediate output Y_i of X_i via $\mathsf{LH}(Y_i) = \mathsf{LH}(C_i)$ and $\mathsf{RH}(Y_i) = G_4(\mathsf{LH}(C_i)) \boxminus \mathsf{RH}(C_i)$. Then run RY on the pairs $\{(X_i, Y_i) \mid i \in \mathcal{I}\}$ to recover G_1, and then recover the round-1 intermediate outputs $Z_1, \ldots, Z_p$ of $X_1, \ldots, X_p$. In addition, observe that $\mathsf{Rev}(C_i)$ is the ciphertext of $\mathsf{Rev}(Z_i)$ under a 3-round Feistel $\overline{\mathsf{F}} = \mathbf{Feistel}[3, N, M, \boxplus]$ of round functions G_2, G_3, G_3 (note that the roles of M and N are now reversed), where for $Z = (A, B) \in \mathbb{Z}_M \times \mathbb{Z}_N$, we write $\mathsf{Rev}(Z)$ to denote the pair $(B, A) \in \mathbb{Z}_N \times \mathbb{Z}_M$. We then run RY on $\mathsf{Rev}(Z_1), \ldots, \mathsf{Rev}(Z_p), \mathsf{Rev}(C_1), \ldots, \mathsf{Rev}(C_p)$ to recover G_2, G_3, G_4.

To amplify the recovery rate, instead of using just one random node x^*, we try μ independent choices of x^*; in our implementation, we pick $\mu = 10$.[11] As analyzed

[11] We note that here $\mu = 10$ means that the attack will iterate up to 10 times, each time with an independent choice of the initial node x^*, until it succeeds in recovering the entire codebook. The expected number of the iterations is often smaller than 10.

Procedure $\mathsf{Restore}(X_1, C_1, \ldots, X_p, C_p, \mathcal{C})$

Pick a node x^* uniformly at random from $\mathcal{C}$
Initialize tables G_1, G_2, G_3, G_4; $G_4(x^*) \leftarrow 0$; $count \leftarrow 0$
Run a breadth-first search on $\mathcal{C}$ from x^* and let T be the corresponding BFS tree
Let $(v_0, \ldots, v_t)$ be the visiting order of the nodes in the BFS above
For ($k = 1$ to t) and $count \le \lfloor 3M/\sqrt{N} \rfloor$ do
  $count \leftarrow count + 1$; Let u be the parent of v_k in T
  $e \leftarrow \{u, v_k\}$; $(w, R) \leftarrow H[e]$
  If $u = w$ then $G_4(v_k) \leftarrow G_4(u) \boxminus R$ else $G_4(v_k) \leftarrow G_4(u) \boxplus R$
$\mathcal{I} \leftarrow \{i \mid G_4(\mathsf{LH}(C_i)) \neq \perp\}$ // Consider the entire set $\{1, \ldots, p\}$ to find $\mathcal{I}$
// Recover the round-3 intermediate outputs for X_i with $i \in \mathcal{I}$
For $i \in \mathcal{I}$ do $Y_i \leftarrow \big(\mathsf{LH}(C_i), \mathsf{RH}(C_i) \boxminus G_4(\mathsf{LH}(C_i))\big)$
Run RY on (X_i, Y_i) for $i \in \mathcal{I}$ to recover G_1 // Here RY attacks domain $\mathbb{Z}_M \times \mathbb{Z}_N$
For $i = 1$ to p do // Recover round-1 intermediate outputs for every X_i
  $(L_0, R_0) \leftarrow X_i$; $L_1 \leftarrow L_0 \boxplus G_1(R_0)$; $R_1 \leftarrow R_0$; $Z_i \leftarrow (L_1, R_1)$
For $i = 1$ to p do $Z'_i \leftarrow \mathsf{Rev}(Z_i)$; $C'_i \leftarrow \mathsf{Rev}(C_i)$
Run RY on (Z'_i, C'_i) to recover (G_2, G_3, G_4) // Now RY attacks domain $\mathbb{Z}_N \times \mathbb{Z}_M$
Return (G_1, G_2, G_3, G_4)

Fig. 12. Procedure $\mathsf{Restore}$ **in the** TF **attack.** Here for $Z = (A, B) \in \mathbb{Z}_M \times \mathbb{Z}_N$, we write $\mathsf{Rev}(Z)$ to denote the pair $(B, A) \in \mathbb{Z}_N \times \mathbb{Z}_M$.

(M, N)	$(2^7, 2^6)$	$(2^7, 2^7)$	$(2^8, 2^7)$	$(2^8, 2^8)$	$(2^9, 2^8)$	$(2^9, 2^9)$	$(10^2, 10^2)$	$(10^3, 10^2)$
$\lvert E^* \rvert$	217 ± 6	203 ± 6	270 ± 6	321 ± 6	802 ± 11	965 ± 11	156 ± 5	1576 ± 16

Table 3. Empirical estimation of $|E^*|$ over 100 trials. The first row indicates the values of M and N. The second row shows the 95% confidence interval of $|E^*|$.

in Section 3.3, we can decide which node yields a correct output by evaluating $3M$ plaintexts on a Feistel network with the recovered round functions, and comparing them with the corresponding ciphertexts.

Analysis. We now analyze the advantage of the TF attack; the key ideas in our analysis are largely from DV's work. We however tighten some of their arguments to improve the bounds.

▷ We begin by estimating $\mathbf{E}[|E^*|]$. A direct generalization of DV's analysis would yield $\mathbf{E}[|E^*|] \approx \frac{p^6}{M^3 N^6}$, which is rather loose. Consider the empirical estimation of $|E^*|$ in Table 3. For $M = N = 100$, the 95% confidence interval of $|E^*|$ is 156 ± 5, but the approximation above suggests that $\mathbf{E}[|E^*|] \approx 400$.

For example, with $M = 1000$ and $N = 100$, empirically the attack would succeed at the first iteration, and thus it only performs a single iteration.

We now provide a tighter analysis. Let W be the number of triangles in $\mathcal{G}$ whose all three nodes are good. From part (a) of Lemma 12, when we enumerate zero-weight triangles in $\mathcal{G}$, most of them will have three good nodes. Thus those triangles will contribute approximately $3W$ distinct good nodes. Note that if $\mathcal{T} = (v_{i,j}, v_{k,\ell}, v_{r,s})$ is a triangle then $\mathcal{T}^* = (v_{j,i}, v_{s,r}, v_{\ell,k})$ is also a triangle, and $\mathsf{weight}(\mathcal{T}^*) = 0 \boxminus \mathsf{weight}(\mathcal{T})$. (See Fig. 9 for an illustration.) Hence if $\mathcal{T}$ is a zero-weight triangle then so is $\mathcal{T}^*$, but these two triangles will produce the same three edges for E^*. Taking into account this duplication, $\mathbf{E}[|E^*|] \approx \frac{3\mathbf{E}[W]}{2}$. From part (b) of Lemma 12,

$$\mathbf{E}[W] \gtrapprox \frac{p^6}{3M^3N^6}\Big(1 - \frac{7}{N} - \frac{14}{M}\Big) \geq \frac{2p^6}{9M^3N^6}$$

for $M \geq N \geq 64$. Hence

$$\mathbf{E}[|E^*|] \gtrapprox \frac{p^6}{3M^3N^6} \gtrapprox \frac{4M}{3} \quad . \tag{5}$$

This lower bound is consistent with Table 3. For example, for $M = N = 100$, we estimate that $\mathbf{E}[|E^*|] \gtrapprox 133$, and recall that empirically, the 95% confidence interval of $|E^*|$ is 156 ± 5.

$\triangleright$ Next, following DV, we model the graph $\mathcal{G}^*$ as a random graph according to the Erdős-Rényi model, in which each of the $\binom{M}{2}$ possible edges will have probability ρ to appear in E^*, independent of other edges. To determine the parameter ρ, note that according to the model above, the expected number of edges in E^* is

$$\binom{M}{2}\rho = \frac{M(M-1)\rho}{2} \quad . \tag{6}$$

From Equation (5) and Equation (6), we have $M\rho = \frac{2\mathbf{E}[|E^*|}{M-1} \gtrapprox \frac{8M}{3}$. From the theory of random graph (see, for example, Chapter 2 of Durrett's book [17]), the graph $\mathcal{G}^*$ will almost surely contain a giant component of size about $(1-c)M$ or bigger, where $c \approx 0.0878$ is the unique solution of the equation $e^{\frac{8}{3}(t-1)} = t$ in the interval $(0, 1)$. In DV's attack, one recovers $G_4(u)$ for every node u in the giant component, but since S *may* contain a few bad nodes, some entries of G_4 that we recover might be incorrect. Instead, we only recover $G_4(u)$ for nodes u in a connected subgraph of the giant component of size $\left\lfloor \frac{3M}{\sqrt{N}} \right\rfloor$. Those nodes are produced by about $\left\lfloor \frac{M}{\sqrt{N}} \right\rfloor$ triangles of zero-weight, and thus from Lemma 12, the chance that the nodes of these triangles are good is at least $1 - \frac{\sqrt{N}}{(M-9)} \cdot \Big(4 + \frac{33N}{(N-2)M} + \frac{39}{M}\Big)$. On the other hand, expectedly, we obtain about

$$\frac{3p}{\sqrt{N}} \geq 3 \cdot 2^{1/3} M^{2/3} \sqrt{N} \geq N(\ln(N) + \ln(2) + 2.7)$$

pairs of plaintext/ciphertext for RY, where the second inequality is due to the fact that $M \geq N \geq 64$. Thus we can run RY to recover G_1 with probability at

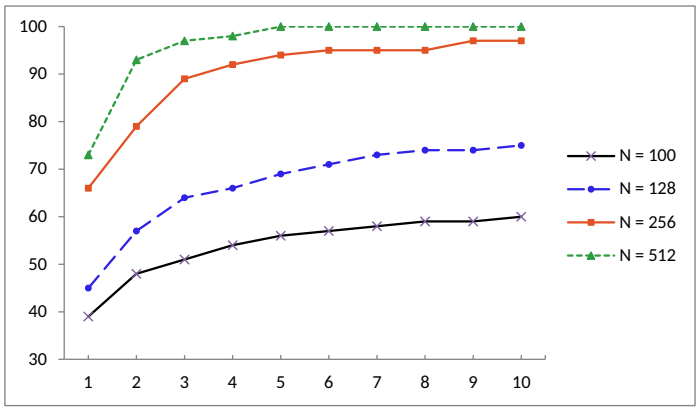

Fig. 13. The performance of TF with respect to μ, on balanced domains $\mathbb{Z}_N \times \mathbb{Z}_N$, over 100 trials. The x-axis indicates the values of μ, and the y-axis shows how many trials, out of 100 ones, that TF can recover the entire codebook.

least $e^{-e^{-2.7}} > 0.935$. We then can run RY with

$$p \geq M(\ln(M) + 5) \geq \max\{N(\ln(N) + 5), M(\ln(M) + \ln(2) + 4.3)\}$$

inputs, and thus can recover (G_2, G_3, G_4) with probability at least $e^{-e^{-4.3}} - e^{-5} > 0.975$. Summing up, if we just try one node x^* then our recovery rate is at least $0.91 - \frac{\sqrt{N}}{(M-9)} \cdot \left(4 + \frac{33N}{(N-2)M} + \frac{39}{M}\right)$. Using μ independent choices of x^* can only improve the success probability. In fact, as illustrated in Fig. 13, empirically, the improvement when we increase μ from 1 to 10 is substantial.

▷ As mentioned in Section 4.1, constructing the differential graph $\mathcal{G}$ takes $O(M^{5/3})$ expected time, and so does enumerating zero-weight triangles. The graph $\mathcal{G}^*$ has M nodes, and expectedly, around $\frac{p^6}{3M^3N^6} \in O(M)$ edges. Thus identifying the connected components of $\mathcal{G}^*$ and doing BFS on its largest component takes $O(|V^*| + \mathbf{E}[|E^*|]) = O(M)$ expected time. Running RY takes $O(p)$ time. Hence the total running time is $O(M^{5/3})$.

Comparison with DV's attack. While our attack is inspired by DV's attack, there are important changes:

- First, as mentioned earlier, compared to DV's notion of differential graphs, we actually use a dual definition, for better recovery rate. Our notion also adds some non-degeneracy requirements, allowing us to find a proof for Lemma 12 and resolve DV's conjecture.
- Next, our attack has a much faster way to enumerate triangles of zero weight, reducing the running time from $O(N^3)$ to $O(N^{5/3})$ in the balanced setting.
- Recall that some zero-weight triangles may contain bad nodes, creating some noise in the attack. DV mentioned that their attack hardly succeeded for $2N^{5/3}$ or more plaintext/ciphertext pairs, and posed an open question to eliminate the noise. To resolve this issue, we introduce the trick of exploring

(M, N)	$(2^7, 2^6)$	$(2^7, 2^7)$	$(2^8, 2^7)$	$(2^8, 2^8)$	$(2^9, 2^8)$	$(2^9, 2^9)$	$(10^2, 10^2)$	$(10^3, 10^2)$
Real	100	100	100	100	100	100	100	100
Ideal	0	0	0	0	0	0	1	0

Table 4. Empirical performance of the LHD attack over 100 trials. The first row indicates the values of M and N. The second row shows how many times, over 100 trials, that LHD correctly outputs 1 in the real world. The last row shows how many times, again over 100 trials, that LHD incorrectly outputs 1 in the ideal world.

---

the giant component from μ random places, and from each place, we stop after visiting $\lceil 3M/\sqrt{N} \rceil$ nodes.

– DV only run RY once to recover (G_1, G_2, G_3), and then derive the round-3 intermediate W_i values of X_i, and use (W_i, C_i) pairs to recover G_4. This works well for DV, as they consider just the balanced setting $M = N$. However, in unbalanced settings, for the first run of RY, we only have $3p/\sqrt{N} \ll M \ln(M)$ inputs. Thus the chance that one can recover (G_1, G_2, G_3) by using one RY call is poor. We therefore only use the first RY call to get G_1, and run RY another time to recover (G_2, G_3, G_4).

5 Experiments

In this section, we empirically evaluate the LHD, SD, and TF attacks.

Benchmarking environment. We implemented our attack in C++ , and ran experiments using 72 threads in a server of dual Intel(R) Xeon(R) CPU E5-2699 v3 2.30GHz CPU and 256 GB RAM. We evaluate our attacks in both balanced and unbalanced settings, and for both binary and decimal domains. Specifically, we consider every $(M, N) \in \{(2^7, 2^6), (2^7, 2^7), (2^8, 2^7), (2^8, 2^8), (2^9, 2^8), (2^9, 2^9), (10^2, 10^2), (10^3, 10^2)\}$. For each choice of (M, N), we let

$$p = \max\Big\{ \lfloor 2^{1/3} M^{2/3} N \rfloor, \lceil M(\ln(M) + 5) \rceil \Big\}$$

as specified in Equation (1).

Evaluating LHD. For each domain $\mathbb{Z}_M \times \mathbb{Z}_N$, we sample p messages uniformly without replacement from $\mathbb{Z}_M \times \mathbb{Z}_N$, and then extract $m = \lceil \frac{p}{N} \cdot \lceil 32N^{1/6} \rceil \rceil$ t-wise right-matching plaintexts, with $t = \lceil \frac{mM}{p} \rceil$. In the real world, we encrypt the plaintexts using the 4-round version of FF3 with the all-zero tweak to produce m ciphertexts. In contrast, the ciphertexts are chosen uniformly without replacement from $\mathbb{Z}_M \times \mathbb{Z}_N$ in the ideal world. The results of our experiments, given in Table 4, show that LHD is nearly perfect, which is much better than our theoretical estimation in Section 3.2. This is not surprising, since our analysis is very conservative.

(M, N)	$(2^7, 2^6)$	$(2^7, 2^7)$	$(2^8, 2^7)$	$(2^8, 2^8)$	$(2^9, 2^8)$	$(2^9, 2^9)$	$(10^2, 10^2)$	$(10^3, 10^2)$
Real	100	100	100	100	100	100	100	100
Ideal	1	11	0	5	0	0	8	0

Table 5. Empirical performance of the LHD attack with aggressive parameters over 100 trials. The first row indicates the values of M and N. The second row shows how many times, over 100 trials, that LHD correctly outputs 1 in the real world. The last row shows how many times, again over 100 trials, that LHD incorrectly outputs 1 in the ideal world.

(M, N)	$(2^7, 2^6)$	$(2^7, 2^7)$	$(2^8, 2^7)$	$(2^8, 2^8)$	$(2^9, 2^8)$	$(2^9, 2^9)$	$(10^2, 10^2)$	$(10^3, 10^2)$
Recover, $\mu = 1$	70	45	75	66	84	73	39	100
Recover, $\mu = 2$	77	57	88	79	93	93	48	100
Recover, $\mu = 3$	81	64	91	89	95	97	51	100
Recover, $\mu = 4$	84	66	94	92	99	98	54	100
Recover, $\mu = 5$	85	69	95	94	100	100	56	100
Recover, $\mu = 6$	86	71	96	95	100	100	57	100
Recover, $\mu = 7$	87	73	97	96	100	100	58	100
Recover, $\mu = 8$	87	74	97	97	100	100	59	100
Recover, $\mu = 9$	88	74	98	97	100	100	59	100
Recover, $\mu = 10$	88	75	98	97	100	100	60	100

Table 6. Empirical performance of the TF attack over 100 trials. The first row indicates the values of M and N. Each subsequent row shows how many trials, over 100 ones, that TF correctly recovers the entire codebook, for the given choice of μ.

We also experiment with more aggressive choices of m to improve performance of LHD. Even if we pick $m = \lceil \frac{p}{N} \cdot \lceil 4N^{1/6} \rceil \rceil$ (meaning an 8x-speedup for LHD), the empirical performance, given in Table 5, is still good.

Evaluating TF. For each domain $\mathbb{Z}_M \times \mathbb{Z}_N$, we sample p messages uniformly without replacement from $\mathbb{Z}_M \times \mathbb{Z}_N$, and generate p ciphertexts using 4-round FF3 with the all-zero tweak. We consider all choices of μ from 1 to 10. The results of our experiments, given in Table 6, are consistent with the theory. For example, with $M = N = 128$ and $\mu = 1$, the attack is supposed to recover the entire codebook with probability around

$$0.91 - \frac{\sqrt{N}}{M-9} \cdot \left(4 + \frac{33N}{(N-2)M} + \frac{39}{M}\right) \approx 47.6\%$$

and in the experiments, 45 out of 100 trials yield the correct codebook. Increasing μ will improve the performance substantially. For example, with $\mu = 10$, the recovery rate goes up to 75%.

(M, N)	$(2^7, 2^6)$	$(2^7, 2^7)$	$(2^8, 2^7)$	$(2^8, 2^8)$	$(2^9, 2^8)$	$(2^9, 2^9)$	$(10^2, 10^2)$	$(10^3, 10^2)$
Have slid pairs	61	65	53	53	38	33	75	27
Survive LHD tests	61	65	53	53	38	33	75	27
Recover	50	39	51	50	38	33	22	27

Table 7. Empirical performance of the SD attack over 100 trials. The first row indicates the values of M and N. The second row shows how many trials, over 100 ones, have at least one slid pair, the third row shows how many of them survive the LHD tests, and the last row shows how many of them can successfully recover the entire codebook.

(M, N)	$(2^7, 2^6)$	$(2^7, 2^7)$	$(2^8, 2^7)$	$(10^2, 10^2)$
No. of candidates	74	72	53	89

Table 8. The total number of survived (possibly false) candidates after using LHD tests in SD, over 100 trials. The first row indicates the values of M and N. The second row shows the total number of survived (possibly false) candidates after using LHD tests in SD, over 100 trials.

Evaluating SD. To save time in evaluating SD, we use the FF3 key to find true candidates and run the LHD tests and the TF attack on them. Table 7 reports the empirical performance of SD over 100 trials, in which we use $\mu = 10$ for the underlying TF attack. The recovery rate is reasonable, ranging from 22% to 51%, and we never miss any true candidate using LHD tests. In addition, we also run the full SD attack on $(M, N) \in \{(2^7, 2^6), (2^7, 2^7), (2^8, 2^7), (10^2, 10^2)\}$ to evaluate the performance of LHD test on false slid-pair candidates. As shown in Table 8, our test is a nearly perfect filtering, leaving on average a single (possibly false) slid-pair candidate in each trial.

Acknowledgments

We thank Adi Shamir and anonymous reviewers of EUROCRYPT 2019 for insightful feedback. Viet Tung Hoang was supported by NSF grants CICI-1738912 and CRII-1755539. Ni Trieu was supported by NSF award #1617197.

References

1. W. Aiello and R. Venkatesan. Foiling birthday attacks in length-doubling transformations - Benes: A non-reversible alternative to Feistel. In U. M. Maurer, editor, *EUROCRYPT'96*, volume 1070 of *LNCS*, pages 307–320. Springer, Heidelberg, May 1996.

2. A. Bar-On, E. Biham, O. Dunkelman, and N. Keller. Efficient slide attacks. *Journal of Cryptology*, 31(3):641–670, July 2018.
3. M. Bellare and V. T. Hoang. Identity-based Format-Preserving Encryption. In *CCS 2017*, pages 1515–1532, 2017.
4. M. Bellare, V. T. Hoang, and S. Tessaro. Message-recovery attacks on feistel-based format preserving encryption. In E. R. Weippl, S. Katzenbeisser, C. Kruegel, A. C. Myers, and S. Halevi, editors, *ACM CCS 16*, pages 444–455. ACM Press, Oct. 2016.
5. M. Bellare, V. T. Hoang, and S. Tessaro. Message-recovery attacks on feistel-based format preserving encryption. In *CCS 2016*, 2016.
6. 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.
7. 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.
8. E. Biham, A. Biryukov, O. Dunkelman, E. Richardson, and A. Shamir. Initial observations on Skipjack: Cryptanalysis of Skipjack-3XOR (invited talk). In S. E. Tavares and H. Meijer, editors, *SAC 1998*, volume 1556 of *LNCS*, pages 362–376. Springer, Heidelberg, Aug. 1999.
9. A. Biryukov, G. Leurent, and L. Perrin. Cryptanalysis of feistel networks with secret round functions. In O. Dunkelman and L. Keliher, editors, *SAC 2015*, volume 9566 of *LNCS*, pages 102–121. Springer, Heidelberg, Aug. 2016.
10. A. Biryukov and D. Wagner. Slide attacks. In L. R. Knudsen, editor, *FSE'99*, volume 1636 of *LNCS*, pages 245–259. Springer, Heidelberg, Mar. 1999.
11. A. Biryukov and D. Wagner. Advanced slide attacks. In B. Preneel, editor, *EUROCRYPT 2000*, volume 1807 of *LNCS*, pages 589–606. Springer, Heidelberg, May 2000.
12. 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.
13. 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`.
14. S. Dara and S. Fluhrer. FNR: Arbitrary length small domain block cipher proposal. In *International Conference on Security, Privacy, and Applied Cryptography Engineering*, pages 146–154. Springer, 2014.
15. F. B. Durak and S. Vaudenay. Breaking and repairing the FF3 format preserving encryption over small domain. In *CRYPTO 2017*, pages 679–707. Springer, 2017.
16. F. B. Durak and S. Vaudenay. Generic round-function-recovery attacks for feistel networks over small domains. In *Applied Cryptography and Network Security*, pages 440–458, 2018.
17. R. Durrett. *Random Graph Dynamics*. Cambridge University Press, 2008.
18. V. T. Hoang, B. Morris, and P. Rogaway. An enciphering scheme based on a card shuffle. In R. Safavi-Naini and R. Canetti, editors, *CRYPTO 2012*, volume 7417 of *LNCS*, pages 1–13. Springer, Heidelberg, Aug. 2012.
19. V. T. Hoang and P. Rogaway. On generalized Feistel networks. In T. Rabin, editor, *CRYPTO 2010*, volume 6223 of *LNCS*, pages 613–630. Springer, Heidelberg, Aug. 2010.
20. V. T. Hoang, S. Tessaro, and N. Trieu. The curse of small domains: New attacks on format-preserving encryption. In *CRYPTO 2018*, pages 221–251. Springer, 2018.

21. U. Mattsson. Format controlling encryption using datatype preserving encryption. Cryptology ePrint Archive, Report 2009/257, 2009. `http://eprint.iacr.org/2009/257`.
22. B. Morris and P. Rogaway. Sometimes-recurse shuffle - almost-random permutations in logarithmic expected time. In P. Q. Nguyen and E. Oswald, editors, *EUROCRYPT 2014*, volume 8441 of *LNCS*, pages 311–326. Springer, Heidelberg, May 2014.
23. R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.
24. 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.
25. J. Patarin. Generic attacks on Feistel schemes. In C. Boyd, editor, *ASIACRYPT 2001*, volume 2248 of *LNCS*, pages 222–238. Springer, Heidelberg, Dec. 2001.
26. T. Ristenpart and S. Yilek. The mix-and-cut shuffle: Small-domain encryption secure against N queries. In R. Canetti and J. A. Garay, editors, *CRYPTO 2013, Part I*, volume 8042 of *LNCS*, pages 392–409. Springer, Heidelberg, Aug. 2013.
27. A. Saltykov. The number of components in a random bipartite graph. *Discrete Mathematics and Applications*, 5(6):515–524, 1995.
28. J. Vance and M. Bellare. Delegatable Feistel-based Format Preserving Encryption mode. Submission to NIST, Nov 2015.

A Attack Notion

In this section, we formalize a notion of *chosen-plaintext codebook-recovery* attacks on FPEs, which captures both our Slide-then-Differential attack, and DV's attack. While the concept of chosen-plaintext codebook-recovery attacks for blockciphers is folklore, for FPE, measuring the advantage of those attacks is somewhat tricky, since (1) the domain might be tiny, and (2) the number of ciphertexts can be more than the domain size. Note that our definition is an attack notion, and thus an FPE scheme meeting this notion might still be insecure for real-world usage.

Chosen-plaintext codebook-recovery attacks. Let F be an FPE scheme. Fix a total ordering on $\mathsf{F.Dom}$ so that we can write $\mathsf{F.Dom} = (M_1, \ldots, M_\ell)$. For an adversary $\mathcal{A}$, define

$$\mathbf{Adv}^{\mathsf{cpa\text{-}cr}}_{\mathsf{F}}(\mathcal{A}) = \Pr[\mathsf{CPA\text{-}Real}_{\mathsf{F}}(\mathcal{A})] - \Pr[\mathsf{CPA\text{-}Ideal}_{\mathsf{F}}(\mathcal{A})]$$

where games $\mathsf{CPA\text{-}Real}_{\mathsf{F}}(\mathcal{A})$ and $\mathsf{CPA\text{-}Ideal}_{\mathsf{F}}(\mathcal{A})$ are defined in Fig. 14. In each game, the adversary is given access to an encryption oracle Enc, and has to output a target tweak T^* and a list of ciphertexts $(C_1, \ldots, C_\ell)$ for messages $(M_1, \ldots, M_\ell)$. The real game returns 1 if every C_i is indeed the ciphertext of M_i and T^* under F. In the ideal game, we instead re-sample $(C_1, \ldots, C_\ell)$ at random, but they are still consistent with the Enc queries that the adversary made. The ideal game returns 1 if every such random C_i is also the ciphertext of M_i and T^* under F.

Games $\mathsf{CPA\text{-}Real}_{\mathsf{F}}(\mathcal{A})$, $\boxed{\mathsf{CPA\text{-}Ideal}_{\mathsf{F}}(\mathcal{A})}$

$K \leftarrow\!\!\$\ \mathsf{F.Keys}$; $(M_1, \ldots, M_\ell) \leftarrow \mathsf{F.Dom}$; $(T^*, C_1, \ldots, C_\ell) \leftarrow\!\!\$\ \mathcal{A}^{\text{Enc}}$

$\Pi \leftarrow\!\!\$\ \text{Perm}(\mathsf{F.Twk}, \mathsf{F.Dom} \mid \text{Map})$

For $i = 1$ to ℓ do $C_i^* \leftarrow \mathsf{F.E}(K, T^*, M_i)$; $\boxed{C_i \leftarrow \Pi(T^*, M_i)}$

Return $(C_1, \ldots, C_\ell) = (C_1^*, \ldots, C_\ell^*)$

Procedure $\text{Enc}(T, M)$

$C \leftarrow \mathsf{F.E}(K, T, M)$; $\text{Map}[T, M] \leftarrow C$; Return C

Fig. 14. Games defining the cpa-cr notion. Game $\mathsf{CPA\text{-}Ideal}_{\mathsf{F}}(\mathcal{A})$ contains the boxed statement, but game $\mathsf{CPA\text{-}Real}_{\mathsf{F}}(\mathcal{A})$ does not. Here $\text{Perm}(\mathcal{T}, \mathcal{M} \mid \text{Map})$ is the set of $\mathcal{T}$-indexed families Π of permutations on $\mathcal{M}$ that are consistent with Map, meaning that $\Pi(T, M) = \text{Map}(T, M)$ for every $(T, M) \in \mathcal{T} \times \mathcal{M}$ such that $\text{Map}[T, M] \neq \perp$.

Discussion. Attacks on FPEs or blockciphers, such as DV's attack, often only report their empirical results on $\Pr[\mathsf{CPA\text{-}Real}_{\mathsf{F}}(\mathcal{A})]$. However, $\Pr[\mathsf{CPA\text{-}Real}_{\mathsf{F}}(\mathcal{A})]$ alone is not enough to indicate the actual attack quality, since an adversary may obtain the entire codebook of T^* via the encryption oracle, and simply output that. To measure the advantage of $\mathcal{A}$, we offset the probability above by $\Pr[\mathsf{CPA\text{-}Ideal}_{\mathsf{F}}(\mathcal{A})]$, the chance that a uniformly random guess of the codebook of T^*, given the plaintext/ciphertext pairs from the encryption oracle, is correct. Blockcipher attacks in cryptanalytic literature often deal with a huge domain size N, say $N = 2^{128}$ and make $q \ll N$ queries. In that case, $\Pr[\mathsf{CPA\text{-}Ideal}_{\mathsf{F}}(\mathcal{A})] \leq \frac{1}{(N-q)!}$, which is extremely small, and it therefore suffices to report $\Pr[\mathsf{CPA\text{-}Real}_{\mathsf{F}}(\mathcal{A})]$ alone.

B Attacking 3-round Feistel-based Blockciphers

In this section, we will generalize DV's known-plaintext attack on 3-round balanced Feistel to the unbalanced case, and name it *Round-wise Yoyo* (RY). We also extract the concrete advantage of the attack from DV's analysis, and tighten some of their arguments to improve the bound in the unbalanced setting. We stress that all key ideas in the attack and analysis are from the work of DV.

In this section, let $\mathsf{F} = \mathbf{Feistel}[3, M, N, \boxplus]$, with round functions F_1, F_2, F_3. We begin with a simple but useful observation of DV.

An observation. For any $\Delta \in \mathbb{Z}_M$, let $\mathsf{Shift}(\mathsf{F}, \Delta)$ denote a 3-round Feistel network $\overline{\mathsf{F}} = \mathbf{Feistel}[3, M, N, \boxplus]$ of round functions $\overline{F}_1, \overline{F}_2$, and $\overline{F}_3$ such that $\overline{F}_1(K, x) = F_1(K, x) \boxminus \Delta$, $\overline{F}_2(K, y) = F_2(K, y \boxplus \Delta)$, and $\overline{F}_3(K, x) = F_3(K, x) \boxplus \Delta$, for any $x \in \mathbb{Z}_N$, $y \in \mathbb{Z}_M$, and any key K. Note that for any choice of Δ, scheme $\overline{\mathsf{F}} = \mathsf{Shift}(\mathsf{F}, \Delta)$ ensures that $\overline{\mathsf{F}}.\mathsf{E}(K, X) = \mathsf{F.E}(K, X)$ for any key K and any $X \in \mathbb{Z}_M \times \mathbb{Z}_N$. Therefore, in a codebook recovery attack against $\mathsf{F.E}(K, \cdot)$, without loss of generality, one can pick a designated point $y^* \in \mathbb{Z}_N$ and *assume* that $F_1(K, y^*) = 0$.

Procedure $\mathsf{RY}(\boldsymbol{X}, \boldsymbol{C})$

Pick $y^* \in \mathbb{Z}_N$ arbitrarily; $G_1(y^*) \leftarrow 0$; *size* $\leftarrow 1$; *loop* $\leftarrow$ true
$S_0 \leftarrow \{(X[i], C[i]) \mid \mathsf{RH}(X[i]) = y^*\}$; $S_1 \leftarrow \emptyset$
While *loop* do // Recover G_1 and G_3
  For each $(X, C) \in S_0$ do $\mathsf{Forward}(X, C)$
  $S_1 \leftarrow \{(\boldsymbol{X}[i], \boldsymbol{C}[i]) \mid G_3(\mathsf{RH}(\boldsymbol{C}[i])) \neq \perp\}$
  For each $(X, C) \in S_1$ do $\mathsf{Backward}(X, C)$
  $S_0 \leftarrow \{(\boldsymbol{X}[i], \boldsymbol{C}[i]) \mid G_1(\mathsf{RH}(\boldsymbol{X}[i])) \neq \perp\}$
  If $|S_0| + |S_1| \leq$ *size* then *loop* $\leftarrow$ false // Terminate at fixed point
  Else *size* $\leftarrow |S_0| + |S_1|$
For each $(X, C) \in S_0$ do // Recover G_2
  $Z \leftarrow G_1(\mathsf{RH}(X)) \boxplus \mathsf{LH}(X)$; $G_2(Z) \leftarrow \mathsf{RH}(C) \boxminus \mathsf{RH}(X)$
Return (G_1, G_2, G_3)

Procedure $\mathsf{Forward}(X, C)$

// Recover intermediate values of X
$X_1 \leftarrow (\mathsf{LH}(X) \boxplus G_1(\mathsf{RH}(X)), \mathsf{RH}(X))$; $X_2 \leftarrow (\mathsf{LH}(X_1), \mathsf{RH}(C))$
$G_3(\mathsf{RH}(C)) \leftarrow \mathsf{LH}(C) \boxminus \mathsf{LH}(X_2)$

Procedure $\mathsf{Backward}(X, C)$

// Recover intermediate values of X
$X_2 \leftarrow (\mathsf{LH}(C) \boxminus G_3(\mathsf{RH}(C)), \mathsf{RH}(C))$; $X_1 \leftarrow (\mathsf{LH}(X_2), \mathsf{RH}(X))$
$G_1(\mathsf{RH}(X)) \leftarrow \mathsf{LH}(X_1) \boxminus \mathsf{LH}(X)$

Fig. 15. The RY **attack.**

THE RY ATTACK. Suppose that we are given a vector $\boldsymbol{X}$ of plaintexts and its corresponding vector $\boldsymbol{C}$ of ciphertexts, where the plaintexts are chosen uniformly without replacement from $\mathbb{Z}_M \times \mathbb{Z}_N$, with $|\boldsymbol{X}| = q$. For $t \in \{1, 2, 3\}$, let $G_t(\cdot)$ denote $F_t(K, \cdot)$ where K is the secret key. Pick an arbitrary designated point $y^* \in \mathbb{Z}_N$. As mentioned above, without loss of generality, we can assume that $G_1(y^*) = 0$. The RY attack will recover the tables of functions G_1, G_2, G_3, instead of just the codebook.

To recover the functions G_1, G_2, G_3, initially the corresponding tables are undefined everywhere, and we set $G_1(y^*) \leftarrow 0$. The attack will enter a loop, trying to update more and more entries of G_1 and G_3 until we reach a fixed point. In each iteration of the loop, we will first use the current table of G_1 to recover some entries of G_3, and then use the newly updated table G_3 to recover more entries of G_1. Thus this loop is essentially a yoyo game [8, 9]. When the loop terminates, we then use the tables of G_1 and G_3 and the pairs $(\boldsymbol{X}, \boldsymbol{C})$ to recover G_2. The code of RY is given in Fig. 15.

IMPLEMENTING RY. To implement the code in Fig. 15 efficiently, instead of maintaining a set S_0 that keeps track of all plaintext/ciphertext pairs $(\boldsymbol{X}[i], \boldsymbol{C}[i])$ such that $G_1(\mathsf{RH}(\boldsymbol{X}[i]))$ is defined, we only need to keep track of *fresh* pairs,

Procedure $\mathsf{RY}(\boldsymbol{X}, \boldsymbol{C})$

For $i = 1$ to $|\boldsymbol{X}|$ do
  $a \leftarrow \mathsf{RH}(\boldsymbol{X}[i])$; $L_0[a].\mathit{list} \leftarrow L_0[a].\mathit{list} \cup \{(\boldsymbol{X}[i], \boldsymbol{C}[i])\}$
  $a \leftarrow \mathsf{RH}(\boldsymbol{C}[i])$; $L_1[a].\mathit{list} \leftarrow L_1[a].\mathit{list} \cup \{(\boldsymbol{X}[i], \boldsymbol{C}[i])\}$
Pick $y^* \in \mathbb{Z}_N$ arbitrarily; $G_1(y^*) \leftarrow 0$; $\mathit{loop} \leftarrow \mathsf{true}$; $S_0 \leftarrow L_0[y^*].\mathit{list}$
While *loop* do // Recover G_1 and G_3
  $S_1 \leftarrow \mathsf{Process}(S_0, L_1, 0)$; $S_0 \leftarrow \mathsf{Process}(S_1, L_0, 1)$
  If $|S_0| + |S_1| = 0$ then $\mathit{loop} \leftarrow \mathsf{false}$ // Terminate at fixed point
$S_0 \leftarrow \{(\boldsymbol{X}[i], \boldsymbol{C}[i]) \mid G_1(\mathsf{RH}(\boldsymbol{X}[i])) \neq \perp\}$
For each $(X, C) \in S_0$ do // Recover G_2
  $Z \leftarrow G_1(\mathsf{RH}(X)) \boxplus \mathsf{LH}(X)$; $G_2(Z) \leftarrow \mathsf{RH}(C) \boxminus \mathsf{RH}(X)$
Return (G_1, G_2, G_3)

Procedure $\mathsf{Process}(S, L, \mathit{sign})$

$S^* \leftarrow \emptyset$
For each $(X, C) \in S$ do
  If $\mathit{sign} = 0$ then $a \leftarrow \mathsf{Forward}(X, C)$ else $a \leftarrow \mathsf{Backward}(X, C)$
  If $\neg(L[a].\mathit{updated})$ then $L[a].\mathit{updated} \leftarrow \mathsf{true}$; $S^* \leftarrow S^* \cup L[a].\mathit{list}$
Return S^*

Procedure $\mathsf{Forward}(X, C)$

// Recover intermediate values of X
$X_1 \leftarrow (\mathsf{LH}(X) \boxplus G_1(\mathsf{RH}(X)), \mathsf{RH}(X))$; $X_2 \leftarrow (\mathsf{LH}(X_1), \mathsf{RH}(C))$
$G_3(\mathsf{RH}(C)) \leftarrow \mathsf{LH}(C) \boxminus \mathsf{LH}(X_2)$
Return $\mathsf{RH}(C)$ // Return the updated entry

Procedure $\mathsf{Backward}(X, C)$

// Recover intermediate values of X
$X_2 \leftarrow (\mathsf{LH}(C) \boxminus G_3(\mathsf{RH}(C)), \mathsf{RH}(C))$; $X_1 \leftarrow (\mathsf{LH}(X_2), \mathsf{RH}(X))$
$G_1(\mathsf{RH}(X)) \leftarrow \mathsf{LH}(X_1) \boxminus \mathsf{LH}(X)$
Return $\mathsf{RH}(X)$ // Return the updated entry

Fig. 16. Implementing RY**.** Sets are implicitly initialized to $\emptyset$, and booleans to false.

meaning that $G_1(\mathsf{RH}(\boldsymbol{X}[i]))$ is just defined in the last yoyo iteration. Maintaining S_1 for G_3 is similar. For fast data access, we store *two* copies of $(\boldsymbol{X}, \boldsymbol{C})$ in arrays L_0 and L_1. Each entry $L_0[a]$, with $a \in \mathbb{Z}_N$, stores all pairs $(\boldsymbol{X}[i], \boldsymbol{C}[i])$ such that $\mathsf{RH}(\boldsymbol{X}[i]) = a$. Likewise, each entry $L_1[a]$ stores all pairs $(\boldsymbol{X}[i], \boldsymbol{C}[i])$ such that $\mathsf{RH}(\boldsymbol{C}[i]) = a$. The code of the implementation is given in Fig. 16.

In the implementation above, preparing the arrays L_0 and L_1 takes $O(q)$ time, and the time to recover G_2 given G_1 and G_3 is also $O(q)$. The yoyo loop basically keeps adding elements to S_0 and S_1, and resetting those sets to $\emptyset$. Since each element $(\boldsymbol{X}[i], \boldsymbol{C}[i])$ appears in S_0 at most once, and the same claim holds for

S_1, the running time of the yoyo loop is also $O(q)$. Hence the total running time is $O(q)$. Below we will give a value of q for RY to achieve a good advantage.

Analysis of RY. Consider the following bipartite graph $\mathcal{G} = (U, V, E)$ of $|U| = N$ left nodes, and $|V| = N$ right nodes. Nodes in U are uniquely labeled with numbers in $\mathbb{Z}_N$, and nodes in V are labeled analogously. The edge set E is constructed as follows. For each pair (X_i, C_i), connect the left node $\mathsf{RH}(X_i) \in U$ with the right node $\mathsf{RH}(C_i) \in V$. What RY does is to look for the connected component in $\mathcal{G}$ containing node $y^* \in U$, and recovers $G_1(u)$ and $G_3(v)$ for every $u \in U$ and $v \in V$ in this connected component. Thus RY can fully recover tables G_1 and G_3 if the graph $\mathcal{G}$ is connected.

To analyze the advantage of RY, we will heuristically model $\mathcal{G}$ as a random bipartite graph of N left nodes and N right nodes, and q edges $(u, v) \leftarrow\!\!\$\, U \times V$ sampled independently. It is known that [27] for large enough N, if $q \geq N \ln(N) + (\ln(2) + \lambda)N$, for any $\lambda \in \mathbb{R}^+$ then the probability that this random graph is connected is close to $e^{-e^{-\lambda}}$.

Next, we determine the sufficient condition for q so that one can fully recover table G_2 once the tables of G_1 and G_3 are already given. We will recover entries $G_2(\mathsf{LH}(V_1)), \ldots, G_2(\mathsf{LH}(V_q))$, where $V_1, \ldots, V_q$ are round-1 intermediate values of $X_1, \ldots, X_q$ respectively. Since $X_1, \ldots, X_q$ are sampled uniformly without replacement from $\mathbb{Z}_M \times \mathbb{Z}_N$ and since the top round of F is a permutation, $V_1, \ldots, V_q$ are also sampled uniformly without replacement from $\mathbb{Z}_M \times \mathbb{Z}_N$. Thus our problem is actually the well-known Coupon-Collector problem: there are M types of coupons and a collector wishes to obtain all M types via buying p coupons of random types.

In the classic setting, for each draw the collector obtains a uniformly random type. In contrast, in our settings, because $Y_1, \ldots, Y_p$ are distinct, each time the collector buys a coupon, its type is slightly biased towards new types that the collector has not yet owned. This means that while the classic bound, stated in Lemma 13 below, continues to apply to our setting, we might need fewer coupons than what is suggested in the classic setting.

Lemma 13 (Coupon collector's problem). *[23, Chapter 3.6] Let $M \geq 1$ be an integer, and let $\delta > 0$ be a real number. Suppose that there are M types of coupon and a collector buys $\tau = \lceil M(\ln(M) + \delta) \rceil$ coupons of truly random types. Then the chance that the collector gets all M types is at least $1 - e^{-\delta}$.*

From Lemma 13, if $q \geq \lceil M(\ln(M) + \delta) \rceil$ then the chance that we can fully recover of G_2 is at least $1 - e^{-\delta}$.

Summing up, if $q \geq \lceil N(\ln(N) + \ln(2) + \lambda) \rceil$ then heuristically, we can fully recover the table of G_1 with probability at least $e^{-e^{-\lambda}}$, for any $\lambda > 0$. If we instead pick $q \geq \max\{\lceil N(\ln(N) + \ln(2) + \lambda) \rceil, \lceil M(\ln(M) + \delta) \rceil\}$ then we can even recover the tables of (G_1, G_2, G_3) with probability around $e^{-e^{-\lambda}} - e^{-\delta}$, for any $\lambda, \delta > 0$.

Experiments. We used the same benchmarking environments in Section 5 to evaluate RY. However unlike other sections, here we also consider parameters M

N	100	128	256	512
Rate	92%	84%	85%	88%

Table 9. Empirical performance of the RY attack over 100 trials, for balanced Feistel. The first row indicates the values of N (which is also M), and the second row indicates the empirical rate, over 100 trials, that RY correctly recovers the entire codebook.

(M, N)	$2^7 \times 2^6$	$2^6 \times 2^7$	$2^8 \times 2^7$	$2^7 \times 2^8$	$2^9 \times 2^8$	$2^8 \times 2^9$	$10^3 \times 10^2$	$10^2 \times 10^3$
Rate	97%	96%	94%	89%	96%	92%	100%	96%

Table 10. Empirical performance of the RY attack over 100 trials, for unbalanced Feistel. The first row indicates the values of M and N, and the second row indicates the empirical rate, over 100 trials, that RY correctly recovers the entire codebook.

---

and N such that $M < N$, since those are needed by the TF attack in Section 4. In particular, we will consider $(M, N) \in \{(2^7, 2^6), (2^7, 2^6), (2^7, 2^7), (2^8, 2^7), (2^7, 2^8), (2^8, 2^8), (2^9, 2^8), (2^8, 2^9), (2^9, 2^9), (10^2, 10^2), (10^3, 10^2), (10^2, 10^3)\}$.

For each domain, we use $q = \max\{\lceil N(\ln(N) + 3)\rceil, \lceil M(\ln(M) + 3)\rceil\}$ plaintexts sampled uniformly without replacement from $\mathbb{Z}_M \times \mathbb{Z}_N$, and generate ciphertexts via a 3-round Feistel network of FF3 round functions on the all-zero tweak. The results of our experiments, given in Tables 9 and 10, are consistent with our thereotical analyses. For example, in balanced domains, empirically there are 84–92 trials out of 100 ones that can successfully recover all round functions, whereas the theory suggests 85% recovery rate.

For unbalanced domains, our empirical recovery rate is better, from 89%–100%. This is however not surprising. Consider, for example, $M = 2^7$ and $N = 2^6$; the experiments suggest an empirical rate of 97% in this case. Note that here $q = 1006 \approx (\ln(2) + 15)N$. Then the chance that we can fully recover (G_1, G_3) is around $e^{-e^{-15}}$, and once we are given (G_1, G_3), we can fully recover G_2 with probability at least 0.95. Hence theoretically, the chance that we can fully recover (G_1, G_2, G_3) is at least 94.99%, which is validated by the experiments with the empirical recovery rate as 97%.

C Deferred Proofs

C.1 Proof of Lemma 1

Let $S' = \{g(V) \mid V \in S^*\}$; note that each element of S' is uniformly distributed over $\mathbb{Z}_M \times \mathbb{Z}_N$, independent of π. Let $\mathcal{I} = \{(i, j) \mid 1 \le i, j \le s\}$, and let $\prec$ be a strict total ordering in $\mathcal{I}$. For any $(i, j) \in \mathcal{I}$, let $\mathrm{Hit}_{i,j}$ denote the event that the ith chain of S and jth chain of S' have at least one slid pair. Then

$$\Pr[P \ge 1] = \Pr\left[\bigcup_{1 \le i,j \le s} \mathrm{Hit}_{i,j}\right] .$$

Clearly, if $s = 1$ then $\Pr[P \geq 1] = \Pr[\mathrm{Hit}_{1,1}]$. If $s \geq 2$ then using Bonferroni's inequality,

$$\Pr\Big[\bigcup_{1\leq i,j\leq s} \mathrm{Hit}_{i,j}\Big] \geq \sum_{1\leq i,j\leq s} \Pr[\mathrm{Hit}_{i,j}] - \sum_{\substack{(i,j),(k,\ell)\in\mathcal{I}\\(i,j)\prec(k,\ell)}} \Pr[\mathrm{Hit}_{i,j}\cap\mathrm{Hit}_{k,\ell}] \ .$$

Fix $(i,j) \in \mathcal{I}$. We will now compute $\Pr[\mathrm{Hit}_{i,j}]$. For each permutation ρ on $\mathbb{Z}_M \times \mathbb{Z}_N$, let G_ρ denote the functional graph of ρ, meaning a directed graph whose nodes are elements of $\mathbb{Z}_M \times \mathbb{Z}_N$, in which a directed edge (u,v) means $v = \rho(u)$. Recall that G_ρ is a collection of disjoint, directed cycles. For two nodes u and v within the same cycle of G_ρ, their *distance* $d_\rho(u,v)$ is the length of the directed path from u to v. Let U and Z be the ith and jth elements of S and S' respectively. Note that $\mathrm{Hit}_{i,j}$ happens if and only if

(1) U and Z stay in the same cycle of G_π that has length at least p, and
(2) either $1 \leq d_\pi(U,Z) \leq p$ or $0 \leq d_\pi(Z,U) \leq p-1$.

For each $t \leq MN$, let C_t be the event that U and Z belong to a t-cycle of G_π. For any $t \leq MN$, the chance that U is in a t-cycle of G_π is

$$\Big(\prod_{i=1}^{t-1} \frac{MN-i}{MN-i+1}\Big)\frac{1}{MN-t+1} = \frac{1}{MN} \ .$$

Since Z is uniformly distributed over $\mathbb{Z}_M \times \mathbb{Z}_N$, independent of U,

$$\Pr[C_t] = \frac{1}{MN}\cdot\frac{t}{MN} = \frac{t}{(MN)^2} \ .$$

Moreover, the probability $\Pr[\mathrm{Hit}_{i,j}]$ can be factored out as follows:

$$\Pr[\mathrm{Hit}_{i,j}] = \Pr\Big[\bigcup_{t=p}^{MN}(\mathrm{Hit}_{i,j}\cap C_t)\Big] = \sum_{t=p}^{MN}\Pr[\mathrm{Hit}_{i,j}\cap C_t] \ ,$$

where the last equality is due to the fact that the events $C_p,\ldots,C_{MN}$ are disjoint. For $p \leq t \leq 2p$, if U and Z stay within a t-cycle of π then $\mathrm{Hit}_{i,j}$ will certainly happens. In other words, $\Pr[\mathrm{Hit}_{i,j}\cap C_t] = \Pr[C_t] = \frac{t}{(MN)^2}$ for every $p \leq t \leq 2p$. For $2p < t \leq MN$, if U and Z stay within a t-cycle of p, then given the position of U, there are exactly $2p$ out of t positions for Z so that $\mathrm{Hit}_{i,j}$ happens. Hence for $2p < t \leq MN$,

$$\Pr[\mathrm{Hit}_{i,j}\cap C_t] = \frac{2p}{t}\Pr[C_t] = \frac{2p}{t}\cdot\frac{t}{(MN)^2} = \frac{2p}{(MN)^2} \ .$$

Summing up,

$$\Pr[\mathrm{Hit}_{i,j}] = \sum_{t=p}^{MN}\Pr[\mathrm{Hit}_{i,j}\cap C_t] = \frac{2p(MN-2p)}{(MN)^2} + \sum_{t=p}^{2p}\frac{t}{(MN)^2}$$

$$= \frac{2p(MN-2p)}{(MN)^2} + \frac{1.5p(p+1)}{(MN)^2} = \delta \ .$$

This justifies part (a).

To justify part (b), fix $(i,j),(k,\ell) \in \mathcal{I}$ such that $(i,j) \prec (k,\ell)$. We now compute $\Pr[\mathrm{Hit}_{i,j} \cap \mathrm{Hit}_{k,\ell}]$. If $i \neq k$ and $j \neq \ell$, then $\mathrm{Hit}_{i,j}$ and $\mathrm{Hit}_{k,\ell}$ are independent, and thus

$$\Pr[\mathrm{Hit}_{i,j} \cap \mathrm{Hit}_{k,\ell}] = (\Pr[\mathrm{Hit}_{i,j}])^2 = \delta^2 \ .$$

Next, suppose that either $i = k$ or $j = \ell$. Without loss of generality, assume that $i = k$. Since $(i,j) \prec (k,\ell)$ and the ordering $\prec$ is strict, $j \neq \ell$. Let U be the i-th element in S, and let Z and Z' be the j-th and ℓ-th element in S' respectively. The event $\mathrm{Hit}_{i,j}$ is the intersection of two events: (i) The event Cycle that U is in a t-cycle of G_π, with $t \geq p$, and (ii) The event D_j that Z falls into the same cycle of G_π with U, and either $1 \leq d_\pi(X,Z) \leq p$ or $0 \leq d_\pi(Z,X) \leq p-1$. Note that

$$\Pr[\mathrm{Cycle}] = \sum_{t=p}^{MN} \frac{1}{MN} = \frac{MN-p+1}{MN} \ .$$

Moreover,

$$\Pr[D_j \mid \mathrm{Cycle}] = \frac{\Pr[\mathrm{Hit}_{i,j}]}{\Pr[\mathrm{Cycle}]} = \frac{MN}{MN-p+1} \cdot \delta \ .$$

Likewise, if we factor $\mathrm{Hit}_{i,\ell}$ as the intersection of Cycle and D_ℓ then

$$\Pr[D_\ell \mid \mathrm{Cycle}] = \frac{MN}{MN-p+1} \cdot \delta \ .$$

On the other hand, note that D_j and D_ℓ are conditionally independent, given Cycle, and thus

$$\begin{aligned}
\Pr[\mathrm{Hit}_{i,j} \cap \mathrm{Hit}_{k,\ell}] &= \Pr[\mathrm{Cycle}] \cdot \Pr[D_j \mid \mathrm{Cycle}] \cdot \Pr[D_\ell \mid \mathrm{Cycle}] \\
&= \frac{MN}{MN-p+1} \cdot \delta^2 \\
&\leq \delta^2 + \frac{p\delta^2}{MN-p} \ .
\end{aligned}$$

Summing up, since there are $\frac{s^2(s^2-1)}{2}$ pairs $(i,j),(k,\ell) \in \mathcal{I}$ such that $(i,j) \prec (k,\ell)$ but only $2s \cdot \frac{s(s-1)}{2} = s^2(s-1)$ such pairs satisfy $i = k$ or $j = \ell$,

$$\begin{aligned}
\Pr[P \geq 1] &\geq s^2\delta - \frac{s^2(s^2-1)\delta^2}{2} - \frac{\delta^2 s^2(s-1)p}{MN-p} \\
&\geq s^2\delta - \frac{s^2(s^2-1)\delta^2}{2} - \delta^2(s-1) \geq s^2\delta - \frac{s^4\delta^2}{2} \geq \frac{s^2\delta}{2} \ ,
\end{aligned}$$

where the second inequality is due to the fact that $s^2 p \leq \frac{MN}{2} \leq MN - p$, and the last inequality is due to the fact that $s^2\delta \leq 1$.

Finally, we describe DV's heuristic estimation of $\Pr[P \geq 1]$. Note that

$$\begin{aligned}\Pr[P \geq 1] &= 1 - \Pr\Big[\bigcap_{1 \leq i,j \leq s} \neg\mathrm{Hit}_{i,j}\Big] \\ &\approx 1 - \prod_{1 \leq i,j \leq s} (1 - \Pr[\mathrm{Hit}_{i,j}]) \\ &\approx 1 - (1 - 2p/MN)^{s^2} \approx 1 - e^{-2s^2 p/MN} \ .\end{aligned}$$

Hence $\Pr[P \geq 1]$ is around $1 - e^{-2s^2 p/MN} \approx 1 - 1/e$. Still, we find that the estimation $\Pr[P \geq 1] \approx 1 - e^{-2s^2 p/MN}$ can be very loose in some settings. For example, for $M = N = 64$, from part (a) $\Pr[P \geq 1] = \frac{11}{32} + \frac{3}{2^{15}} \approx 0.344$ which is considerably smaller than $1 - 1/e \approx 0.632$. In contrast, the estimation $\Pr[P \geq 1] \approx 1 - e^{-2s^2 p/MN}$ provides a much tighter bound $1 - 1/\sqrt{e} \approx 0.393$.

C.2 Proof of Lemma 2

For each permutation ρ on $\mathbb{Z}_M \times \mathbb{Z}_N$, let G_ρ denote the functional graph of ρ, meaning a directed graph whose nodes are elements of $\mathbb{Z}_M \times \mathbb{Z}_N$, in which a directed edge (u, v) means $v = \rho(u)$. Recall that G_ρ is a collection of disjoint, directed cycles. For two nodes u and v within the same cycle of G_ρ, their *distance* $d_\rho(u, v)$ is the length of the directed path from u to v. Let $S' = \{g(V) \mid V \in S^*\}$; note that each element of S' is uniformly distributed over $\mathbb{Z}_M \times \mathbb{Z}_N$, independent of π. For any $1 \leq i, j \leq s$, let $Q_{i,j}$ denote the random variable for the number of slid pairs that the ith chain of S and jth chain of S' form. Then

$$P \leq \sum_{1 \leq i,j \leq s} Q_{i,j}$$

and taking expectation of both sides gives us

$$\mathbf{E}[P] \leq \sum_{1 \leq i,j \leq s} \mathbf{E}[Q_{i,j}] \ .$$

Fix $i, j \leq s$, and let U and Z be the ith and j elements of S and S' respectively. It suffices to prove that $\mathbf{E}[Q_{i,j}] \leq \frac{2p}{MN}$. Note that $Q_{i,j} \geq 1$ if and only if (1) Z and U stay in the same cycle of G_π that has length at least p, and (2) either $1 \leq d_\pi(U, Z) \leq p$ or $0 \leq d_\pi(Z, U) \leq p-1$. In addition, $Q_{i,j} = 2$ if and only if the cycle above has length $p + \ell$, with $0 \leq \ell \leq p - 1$, and $\ell - 1 \leq d_\pi(Z, U) \leq p - 1$. For each $k \leq MN$, let B_k be the indicator random variable for the event that U and Z belong to a k-cycle of G_π. Then

$$\mathbf{E}[Q_{i,j}] = \mathbf{E}\Big[Q_{i,j} \cdot \sum_{k=p}^{MN} B_k\Big] = \sum_{k=p}^{MN} \mathbf{E}[Q_{i,j} B_k] \ .$$

Fix $k \in \{p, \ldots, MN\}$. It suffices to show that $\mathbf{E}[Q_{i,j}B_k] \leq \frac{2p}{(MN)^2}$. Note that the chance that U stays within a k-cycle of G_π is exactly:

$$\Big(\prod_{i=1}^{k-1} \frac{MN-i}{MN-i+1}\Big)\frac{1}{MN-k+1} = \frac{1}{MN} \ .$$

We consider the following cases.

Case 1: $k = p$. If U and Z belong to a p-cycle of G_π then $Q_{i,j} = 2$. Since Z is chosen uniformly from $\mathbb{Z}_M \times \mathbb{Z}_N$, independent of S and π,

$$\mathbf{E}[Q_{i,j}B_p] = 2 \cdot \frac{1}{MN} \cdot \frac{p}{MN} = \frac{2p}{(MN)^2} \ .$$

Case 2: $p+1 \leq k \leq 2p-1$. If U belongs to a k-cycle of G_π then there are $2p-1-k$ positions for Z such that $Q_{i,j} = 2$, and there are $2k+1-2p$ positions for Z such that $Q_{i,j} = 1$. Hence

$$\mathbf{E}[Q_{i,j}B_k] = \frac{1}{NM} \cdot \frac{(2k+1-2p)+2(2p-1-k)}{MN} = \frac{2p-1}{(MN)^2} \leq \frac{2p}{(MN)^2} \ .$$

Case 3: $k \geq 2p$. If U belongs to a k-cycle of G_π then there are exactly $2p$ positions for Z such that $Q_{i,j} = 1$, and $Q_{i,j} = 0$ otherwise. Hence

$$\mathbf{E}[Q_{i,j}B_k] = \frac{2p}{(MN)^2} \ .$$

Hence in every case, we have $\mathbf{E}[Q_{i,j}B_k] \leq \frac{2p}{(MN)^2}$ as claimed. This concludes the proof.

C.3 Proof of Lemma 3

Let X_i and X'_i be the round-i intermediate values of X and X' respectively. Let $(G_1, \ldots, G_4)$ be the functions specified by the key K. From part (a) of Lemma 4, $\mathsf{RH}(X_2)$ and $\mathsf{RH}(X'_2)$ are uniformly and independently distributed over $\mathbb{Z}_N$. We consider the following two cases.

Case 1: $\mathsf{RH}(X_2) = \mathsf{RH}(X'_2)$, which happens with probability $1/N$. Then using part (b) of Lemma 4, in this case we always have $\mathsf{LH}(C) \boxminus \mathsf{LH}(C') = \mathsf{LH}(X) \boxminus \mathsf{LH}(X')$.

Case 2: $\mathsf{RH}(X_2) \neq \mathsf{RH}(X'_2)$. This case happens with probability $1 - 1/N$. Using part (c) of Lemma 4, in this case, the conditional probability that $\mathsf{LH}(C) \boxminus \mathsf{LH}(C') = \mathsf{LH}(X) \boxminus \mathsf{LH}(X')$ is $1/M$.

Combining the two cases,

$$\begin{aligned}\Pr[\mathsf{LH}(C) \boxminus \mathsf{LH}(C') = \mathsf{LH}(X) \boxminus \mathsf{LH}(X')] &= \frac{1}{N} + \Big(1 - \frac{1}{N}\Big)\frac{1}{M} \\ &= \frac{M+N-1}{MN} \ .\end{aligned}$$

This concludes the proof.

C.4 Proof of Lemma 4

Let $(G_1, \ldots, G_4)$ be the functions specified by the key K. First consider part(a). Since $X \neq X'$, we must have $\mathsf{LH}(X) \neq \mathsf{LH}(X')$. On the one hand,

$$\mathsf{LH}(X_1) = G_1(\mathsf{RH}(X)) \boxplus \mathsf{LH}(X) \neq G_1(\mathsf{RH}(X')) \boxplus \mathsf{LH}(X') = \mathsf{LH}(X_1') \ .$$

On the other hand, recall that

$$\begin{aligned} \mathsf{RH}(X_2) &= G_2(\mathsf{LH}(X_1) \boxplus \mathsf{RH}(X_1)), \text{ and} \\ \mathsf{RH}(X_2')) &= G_2(\mathsf{LH}(X_1') \boxplus \mathsf{RH}(X_2') \ . \end{aligned}$$

Since G_2 is a truly random function, independent of X_1 and X_1', the random variables X_2 and X_2' are uniformly and independently distributed over $\mathbb{Z}_N$.

Next consider part (b). Then

$$\begin{aligned} &\mathsf{LH}(C) \boxminus \mathsf{LH}(C') = \mathsf{LH}(X_3) \boxminus \mathsf{LH}(X_3') \\ &= \Big(\mathsf{LH}(X_2) \boxplus G_3(\mathsf{RH}(X_2))\Big) \boxminus \Big(\mathsf{LH}(X_2') \boxplus G_3(\mathsf{RH}(X_2'))\Big) \\ &= \mathsf{LH}(X_2) \boxminus \mathsf{LH}(X_2') \ , \end{aligned}$$

where the last equality is due to the fact that $\mathsf{RH}(X_2) = \mathsf{RH}(X_2')$. On the other hand, since $\mathsf{LH}(X_2) = \mathsf{LH}(X_1)$ and $\mathsf{LH}(X_2') = \mathsf{LH}(X_1')$,

$$\begin{aligned} &\mathsf{LH}(X_2) \boxminus \mathsf{LH}(X_2') = \mathsf{LH}(X_1) \boxminus \mathsf{LH}(X_1') \\ &= \Big(\mathsf{LH}(X) \boxplus G_1(\mathsf{RH}(X))\Big) \boxminus \Big(\mathsf{LH}(X') \boxplus G_1(\mathsf{RH}(X'))\Big) \\ &= \mathsf{LH}(X) \boxminus \mathsf{LH}(X') \ , \end{aligned}$$

where the last equality is due to the fact that $\mathsf{RH}(X) = \mathsf{RH}(X')$. Hence we have $\mathsf{LH}(C) \boxminus \mathsf{LH}(C') = \mathsf{LH}(X) \boxminus \mathsf{LH}(X')$.

Finally consider part (b). Then

$$\begin{aligned} &\mathsf{LH}(C) \boxminus \mathsf{LH}(C') = \mathsf{LH}(X_3) \boxminus \mathsf{LH}(X_3') \\ &= \Big(\mathsf{LH}(X_2) \boxplus G_3(\mathsf{RH}(X_2))\Big) \boxminus \Big(\mathsf{LH}(X_2') \boxplus G_3(\mathsf{RH}(X_2'))\Big) \ . \end{aligned}$$

Since $\mathsf{RH}(X_2) \neq \mathsf{RH}(X_2')$ and $G_3 : \mathbb{Z}_N \to \mathbb{Z}_M$ is a truly random function that is independent of $\mathsf{RH}(X_2)$ and $\mathsf{RH}(X_2')$, the random variable $\mathsf{LH}(C) \boxminus \mathsf{LH}(C')$ is uniformly distributed over $\mathbb{Z}_M$, and thus in this case, the conditional probability that $\mathsf{LH}(C) \boxminus \mathsf{LH}(C') = \mathsf{LH}(X) \boxminus \mathsf{LH}(X')$ is $1/M$.

C.5 Proof of Lemma 5

Let

$$\delta = \Delta - \frac{N}{MN-1} = \lambda \cdot \Big(\frac{M+N-1}{MN} - \frac{N}{MN-1}\Big) \geq \frac{\lambda(M-2)}{MN-1} \ .$$

First, recall that

$$\mathbf{E}[V] = \frac{N}{MN-1} \cdot \mathit{size} \ .$$

Hence

$$\begin{aligned}\Pr\left[V \geq \Delta \cdot \mathit{size}\right] &= \Pr\left[V \geq \mathbf{E}[V] + \delta \cdot \mathit{size}\right] \\ &\leq \frac{\mathbf{Var}[V]}{\mathbf{Var}[V] + \delta^2 \cdot (\mathit{size})^2} \leq \frac{\mathbf{Var}[V]}{\delta^2 \cdot (\mathit{size})^2} \ , \end{aligned} \tag{7}$$

where the first inequality is due to one-sided Chebyshev's inequality. Next, we will give a lower bound of *size*. Partition $X_1, \ldots, X_m$ to groups $P_1, \ldots, P_d$ such that X_i and X_j belong to the same partition only if they have the same right segment. Since $X_1, \ldots, X_m$ are t-wise right-matching, we have $d \leq t$. For each $j \leq d$, let m_j be the size of P_j. Then

$$\begin{aligned}\mathit{size} &= \sum_{j=1}^{d} \frac{m_j(m_j-1)}{2} = \frac{1}{2}\Big(\sum_{j=1}^{d} m_j^2\Big) - \frac{1}{2}\Big(\sum_{j=1}^{d} m_j\Big) \\ &= \frac{1}{2}\Big(\sum_{j=1}^{d} m_j^2\Big) - \frac{m}{2} \geq \frac{1}{2d}\Big(\sum_{j=1}^{d} m_j\Big)^2 - \frac{m}{2} = \frac{m^2}{2d} - \frac{m}{2} \geq \frac{m^2}{2t} - \frac{m}{2} \end{aligned} \tag{8}$$

where the first inequality is due to Cauchy-Schwartz inequality. From Equation (7) and Equation (8), it suffices to show that

$$\mathbf{Var}[V] \leq \mathit{size} \cdot \frac{N}{MN-1} + (\mathit{size})^2 \frac{N^2}{(MN-2)(MN-1)^2} \ .$$

In order to do that, we will factor V as

$$V = \sum_{(i,j)\in\mathcal{D}} B_{i,j}$$

where $B_{i,j}$ is the indicator random variable for the event $\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j) = \mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j)$, and $\mathcal{D}$ is the set of all pairs (i,j) such that $1 \leq i < j \leq m$, and $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$. Hence

$$\mathbf{Var}[V] = \sum_{(i,j),(k,\ell)\in\mathcal{D}} \mathbf{Cov}(B_{i,j}, B_{k,\ell}) \ . \tag{9}$$

If $i = k$ and $j = \ell$ then

$$\mathbf{Cov}(B_{i,j}, B_{k,\ell}) = \mathbf{Var}[B_{i,j}] \leq \Pr[B_{i,j} = 1] = \frac{N}{MN-1} \ , \tag{10}$$

where the inequality is due to the fact that $B_{i,j} \in \{0,1\}$. If $(i,j) \neq (k,\ell)$ then we claim that

$$\mathbf{Cov}(B_{i,j}, B_{k,\ell}) \leq \frac{N^2}{(MN-2)(MN-1)^2} \ . \tag{11}$$

Combining Equation (10) and Equation (11) gives us the desired bound on $\mathbf{Var}[V]$. To prove Equation (11), note that since $\Pr[B_{i,j}] = \Pr[B_{k,\ell}] = \frac{N}{MN-1}$,

$$\begin{aligned}\mathbf{Cov}(B_{i,j}, B_{k,\ell}) &= \Pr[B_{i,j} \cdot B_{k,\ell} = 1] - \Pr[B_{i,j} = 1] \cdot \Pr[B_{k,\ell} = 1] \\ &= \Pr[B_{i,j} \cdot B_{k,\ell} = 1] - \frac{N^2}{(MN-1)^2} \ . \end{aligned} \tag{12}$$

We now consider the following cases.

Case 1: $i = k$ or $j = \ell$, but $(i, j) \neq (k, \ell)$. Without loss of generality, assume that $i = k$. Note that in this case, X_i, X_j, X_ℓ have the same right segments, and thus their left segments are distinct. If $\mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j) = \mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j)$ and $\mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_\ell) = \mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_\ell)$ then given X_i, X_j, X_ℓ, and C_i, there are N possible values for C_j and also N possible values for C_ℓ, and those N^2 pairs are equally likely. Hence

$$\Pr[B_{i,j} \cdot B_{k,\ell} = 1] = \frac{N^2}{(MN-1)(MN-2)} \ .$$

Thus from Equation (12),

$$\mathbf{Cov}[B_{i,j}, B_{k,\ell}] = \frac{N^2}{(MN-1)(MN-2)} - \frac{N^2}{(MN-1)^2} = \frac{N^2}{(MN-2)(MN-1)^2} \ .$$

Case 2: $i \neq k$ and $j \neq \ell$ and $\mathsf{RH}(X_i) \neq \mathsf{RH}(X_k)$. Note that $\mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j)$ and $\mathsf{LH}(X_k) \boxminus \mathsf{LH}(X_\ell)$ are independently and uniformly distributed over $\mathbb{Z}_M \backslash \{0\}$, and they are independent of C_i, C_j, C_k, C_ℓ. Hence

$$\Pr[B_{i,j} \cdot B_{k,\ell} = 1] = \frac{\Pr[\overline{R_1} \cap \overline{R_2}]}{(M-1)^2} = \frac{1 - \Pr[R_1 \cup R_2]}{(M-1)^2} \ , \tag{13}$$

where $R_1$is the event $\mathsf{LH}(C_i) = \mathsf{LH}(C_j)$, and R_2 is the event $\mathsf{LH}(C_k) = \mathsf{LH}(C_\ell)$. Note that $\Pr[R_1] = \Pr[R_2] = \frac{N-1}{MN-1}$. By Principle of Inclusion and Exclusion,

$$\begin{aligned}\Pr[R_1 \cup R_2] &= \Pr[R_1] + \Pr[R_2] - \Pr[R_1 \cap R_2] \\ &= \frac{2(N-1)}{MN-1} - \Pr[R_1] \cdot \Pr[R_2 \mid R_1] \\ &= \frac{2(N-1)}{MN-1} - \frac{N-1}{MN-1} \cdot \Pr[R_2 \mid R_1] \ . \end{aligned} \tag{14}$$

We now compute $\Pr[R_2 \mid R_1]$. Fix C_i and C_j such that $\mathsf{LH}(C_i) = \mathsf{LH}(C_j)$ and $C_i \neq C_j$. For any $s \in \mathbb{Z}_M \backslash \{\mathsf{LH}(C_i)\}$, if $\mathsf{LH}(C_k) = \mathsf{LH}(C_\ell) = s$ then there are $N(N-1)$ possible values for (C_k, C_ℓ). If $\mathsf{LH}(C_k) = \mathsf{LH}(C_\ell) = \mathsf{LH}(C_i)$ then there are only $(N-2)(N-3)$ possible values for (C_k, C_ℓ). Hence there are totally $(N-2)(N-3) + (M-1)N(N-1)$ possible values for (C_k, C_ℓ), and these values are equally likely. Thus

$$\Pr[R_2 \mid R_1] = \frac{(N-2)(N-3) + (M-1)N(N-1)}{(MN-2)(MN-3)} \ . \tag{15}$$

Combining Equations (12), (13), (14), (15), and using some algebraic manipulations,

$$\mathbf{Cov}(B_{i,j}, B_{k,\ell}) = \frac{N(N-1)}{(MN-1)^2(MN-2)(MN-3)} \leq \frac{N^2}{(MN-1)^2(MN-2)} \ .$$

Case 3: $i \neq k$ and $j \neq \ell$, but $\mathsf{RH}(X_i) = \mathsf{RH}(X_k)$. In other words, X_i, X_j, X_k, X_ℓ are distinct but have the same right segment. We now show that

$$\begin{aligned} &\Pr[B_{k,\ell} = 1 \mid B_{i,j} = 1] \\ &\leq \frac{N}{MN-3} - \frac{2N}{(MN-3)(MN-2)} + \frac{2}{(MN-3)(MN-2)(M-3)} \ , \end{aligned} \tag{16}$$

and thus from Equation (12) with some algebraic manipulations,

$$\mathbf{Cov}(B_{i,j}, B_{k,\ell}) \leq \frac{N^2}{(MN-1)^2(MN-2)} \ .$$

To justify Equation (16), fix distinct $X_i, X_j, X_k \in \mathbb{Z}_M \times \mathbb{Z}_N$ of the same right segment, and fix distinct $C_i, C_j \in \mathbb{Z}_M \times \mathbb{Z}_N$ such that $\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j) = \mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_k)$. Sample X_ℓ uniformly from $\mathbb{Z}_M \times \mathbb{Z}_N$, subject to the condition that $X_\ell \notin \{X_i, X_j, X_k\}$ and X_ℓ is of the same right segment as X_i, X_j, X_k. Also sample C_k and C_ℓ uniformly without replacement from $\mathbb{Z}_M \times \mathbb{Z}_N \backslash \{C_i, C_j\}$. Let Bad be the event $\mathsf{LH}(X_k) \boxminus \mathsf{LH}(X_\ell) \in \{\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j), \mathsf{LH}(C_j) \boxminus \mathsf{LH}(C_i)\}$. Then $\Pr[\mathsf{Bad}] \leq \frac{2}{M-3}$. We consider the following sub-cases.

Case 3.1: Bad does not happen. Let R be the event that $\mathsf{LH}(X_k) \boxminus \mathsf{LH}(X_\ell) \in \{\mathsf{LH}(C_k) \boxminus \mathsf{LH}(C_i), \mathsf{LH}(C_k) \boxminus \mathsf{LH}(C_j)\}$. Then in this sub-case, $\Pr[R] = \frac{2N}{MN-2}$. Moreover,

$$\begin{aligned} \Pr[B_{k,\ell} = 1 \mid \overline{R}] &= \frac{N}{MN-3} \ , \text{ and} \\ \Pr[B_{k,\ell} = 1 \mid R] &= \frac{N-1}{MN-3} \ . \end{aligned}$$

Summing up, in this sub-case,

$$\begin{aligned} \Pr[B_{k,\ell} = 1] &= \frac{N}{MN-3} \cdot \Pr[\overline{R}] + \frac{N-1}{MN-3} \cdot \Pr[R] \\ &= \frac{N}{MN-3} - \frac{2N}{(MN-3)(MN-2)} \ . \end{aligned}$$

Case 3.2: Bad happens. Let R be the event that $\mathsf{LH}(X_k) \boxminus \mathsf{LH}(X_\ell) \in \{\mathsf{LH}(C_k) \boxminus \mathsf{LH}(C_i), \mathsf{LH}(C_k) \boxminus \mathsf{LH}(C_j)\}$. Then in this sub-case, $\Pr[R] \geq \frac{2N-2}{MN-2}$. Moreover,

$$\Pr[B_{k,\ell} = 1 \mid \overline{R}] = \frac{N}{MN-3} \ , \text{ and}$$

$$\Pr[B_{k,\ell} = 1 \mid R] = \frac{N-1}{MN-3} \ .$$

Summing up, in this sub-case,

$$\begin{aligned}\Pr[B_{k,\ell} = 1] &= \frac{N}{MN-3} \cdot \Pr[\overline{R}] + \frac{N-1}{MN-3} \cdot \Pr[R] \\ &= \frac{N}{MN-3} - \frac{1}{MN-3} \cdot \Pr[R] \\ &\leq \frac{N}{MN-3} - \frac{2N-2}{(MN-3)(MN-2)} \ .\end{aligned}$$

Combining both Cases 3.1 and 3.2 above,

$$\begin{aligned}&\Pr[B_{k,\ell} = 1 \mid B_{i,j} = 1] \\ &\leq \frac{N}{MN-3} - \frac{2N}{(MN-3)(MN-2)} + \frac{2}{(MN-3)(MN-2)} \Pr[\mathsf{Bad}] \\ &\leq \frac{N}{MN-3} - \frac{2N}{(MN-3)(MN-2)} + \frac{2}{(MN-3)(MN-2)(M-3)}\end{aligned}$$

as claimed.

C.6 Proof of Lemma 6

Let

$$\delta = \frac{M+N-1}{MN} - \Delta = (1-\lambda)\Big(\frac{M+N-1}{MN} - \frac{N}{MN-1}\Big) \geq \frac{(1-\lambda)(M-2)}{MN} \ .$$

First, recall that

$$\mathbf{E}[V] = \frac{M+N-1}{MN} \cdot \mathit{size} \ .$$

Hence

$$\begin{aligned}\Pr\left[V \leq \Delta \cdot \mathit{size}\right] &= \Pr\left[V \leq \mathbf{E}[V] - \delta \cdot \mathit{size}\right] \\ &\leq \frac{\mathbf{Var}[V]}{\mathbf{Var}[V] + \delta^2 \cdot (\mathit{size})^2} \leq \frac{\mathbf{Var}[V]}{\delta^2 \cdot (\mathit{size})^2} \ , \end{aligned} \tag{17}$$

where the first inequality is due to one-sided Chebyshev's inequality. Next, we will give a lower bound of *size*. Partition $X_1, \ldots, X_m$ to groups $P_1, \ldots, P_d$ such that X_i and X_j belong to the same partition only if they have the same right segment. Since $X_1, \ldots, X_m$ are t-wise right-matching, we have $d \leq t$. For each $j \leq d$, let m_j be the size of P_j. Then

$$\mathit{size} = \sum_{j=1}^{d} \frac{m_j(m_j-1)}{2} = \frac{1}{2}\Big(\sum_{j=1}^{d} m_j^2\Big) - \frac{1}{2}\Big(\sum_{j=1}^{d} m_j\Big)$$

$$= \frac{1}{2}\Big(\sum_{j=1}^{d} m_j^2\Big) - \frac{m}{2} \geq \frac{1}{2d}\Big(\sum_{j=1}^{d} m_j\Big)^2 - \frac{m}{2} = \frac{m^2}{2d} - \frac{m}{2} \geq \frac{m^2}{2t} - \frac{m}{2} \quad (18)$$

where the first inequality is due to Cauchy-Schwartz inequality. From Equation (7) and Equation (18), it suffices to show that

$$\mathbf{Var}[V] \leq \frac{(M+N-1)\cdot \mathit{size}}{MN} + \frac{6.2(M-1)(N-1)\cdot(\mathit{size})^2}{M^2N^3} + \frac{2\sqrt{2}\cdot(\mathit{size})^{1.5}}{MN} \ .$$

In order to do that, we will factor V as

$$V = \sum_{(i,j)\in\mathcal{D}} B_{i,j}$$

where $B_{i,j}$ is the indicator random variable for the event $\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j) = \mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j)$, and $\mathcal{D}$ is the set of all pairs (i,j) such that $1 \leq i < j \leq m$, and $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$. Hence

$$\mathbf{Var}[V] = \sum_{(i,j),(k,\ell)\in\mathcal{D}} \mathbf{Cov}(B_{i,j}, B_{k,\ell}) \ . \quad (19)$$

If $i = k$ and $j = \ell$ then

$$\mathbf{Cov}(B_{i,j}, B_{k,\ell}) = \mathbf{Var}[B_{i,j}] \leq \Pr[B_{i,j} = 1] = \frac{M+N-1}{MN} \ , \quad (20)$$

where the inequality is due to the fact that $B_{i,j} \in \{0,1\}$, and the last equality is due to Lemma 3. We claim that

(i) Let *pairs* be the number of pairs $(i,j),(i,\ell) \in \mathcal{D}$ such that either $\{i,j\} \cap \{k,\ell\} \neq \emptyset$, but $(i,j) \neq (k,\ell)$. Then

$$\mathit{pairs} \leq 2\sqrt{2}\cdot(\mathit{size})^{1.5} \ . \quad (21)$$

(ii) If $\{i,j\} \cap \{k,\ell\} \neq \emptyset$ but $(i,j) \neq (k,\ell)$ then

$$0 < \mathbf{Cov}(B_{i,j}, B_{k,\ell}) = \frac{MN - M - N + 2}{(MN)^2} < \frac{1}{MN} \ . \quad (22)$$

(iii) If $\{i,j\} \cap \{k,\ell\} = \emptyset$ and $\mathsf{RH}(X_i) = \mathsf{RH}(X_k)$ then

$$\mathbf{Cov}(B_{i,j}, B_{k,\ell}) = \frac{2(M-1)(N-1)}{M^2N^3} \leq \frac{6.2\cdot(M-1)(N-1)}{M^2N^3} \ . \quad (23)$$

(iv) If $\{i,j\} \cap \{k,\ell\} = \emptyset$ and $\mathsf{RH}(X_i) \neq \mathsf{RH}(X_k)$ then

$$\mathbf{Cov}(B_{i,j}, B_{k,\ell}) \leq \frac{6.2\cdot(M-1)(N-1)}{M^2N^3} \ . \quad (24)$$

These claims will be justified later. Combining Equations (20), (21), (22), (23), and (24), we obtain Equation (19).

Justifying Equation (21). If $\{i,j\} \cap \{k,\ell\} \neq \emptyset$, but $(i,j) \neq (k,\ell)$ then X_i, X_j, X_k, X_ℓ must belong to one partition P_s. Let $\mathcal{I}_s$ be the set of indices r such that $X_r \in P_s$. Note that there is a one-to-one correspondence between such a tuple (i,j,k,ℓ) with a tuple (r_1,r_2,r_3) such that $r_1, r_2, r_3 \in \mathcal{I}_s$ are distinct:

(i) Given (i,j,k,ℓ), we can construct (r_1,r_2,r_3) by deleting a duplicate number in (i,j,k,ℓ).
(ii) Given (r_1,r_2,r_3), we can reconstruct (i,j,k,ℓ) via $i = \min\{r_1,r_2\}, j = \max\{r_1,r_2\}, k = \min\{r_2,r_3\}$, and $\ell = \max\{r_2,r_3\}$.

Let $\mu = \max\{m_1,\ldots,m_d\}$. Then

$$\mathit{pairs} = \sum_{t=1}^{d} m_t(m_t-1)(m_t-2) \leq (\mu-2)\sum_{t=1}^{d} m_t(m_t-1) = 2(\mu-2)\mathit{size} \ .$$

On the other hand,

$$\mathit{size} = \sum_{t=1}^{d} \frac{m_t(m_t-1)}{2} \geq \frac{(\mu-1)\mu}{2} \geq \frac{(\mu-2)^2}{2} \ .$$

Thus

$$\mathit{pairs} \leq 2(\mu-2)\mathit{size} \leq 2\sqrt{2}\cdot \mathit{size}^{1.5}$$

as claimed.

Justifying Equation (22). Without loss of generality, assume that $i = k$. For any $r \leq 4$, let X_i^r be the intermediate round-r output of X_i. Let $(G_1,\ldots,G_4)$ be the functions specified by the key of $\overline{\mathsf{F}}$. Since $X_i \neq X_j$ and $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$, we must have $\mathsf{LH}(X_i) \neq \mathsf{LH}(X_j)$. Hence

$$\begin{aligned}\mathsf{LH}(X_i^1) &= \mathsf{LH}(X_i) \boxplus G_1(\mathsf{RH}(X_i)) \\ &\neq \mathsf{LH}(X_j) \boxplus G_1(\mathsf{RH}(X_j)) = \mathsf{LH}(X_j^1) \ .\end{aligned}$$

Repeating this argument, we conclude that the left segments of X_i^1, X_j^1, X_ℓ^1 are distinct. We consider the following cases.

Case 1: $\mathsf{RH}(X_i^2) = \mathsf{RH}(X_j^2)$ and $\mathsf{RH}(X_i^2) = \mathsf{RH}(X_\ell^2)$, meaning that

$$\begin{aligned}\mathsf{RH}(X_i^1) \boxplus G_2(\mathsf{LH}(X_i^1)) &= \mathsf{RH}(X_j^1) \boxplus G_2(\mathsf{LH}(X_j^1)) \\ \mathsf{RH}(X_i^1) \boxplus G_2(\mathsf{LH}(X_i^1)) &= \mathsf{RH}(X_\ell^1) \boxplus G_2(\mathsf{LH}(X_\ell^1)) \ .\end{aligned}$$

Since $\mathsf{LH}(X_i^1), \mathsf{LH}(X_j^1), \mathsf{LH}(X_\ell^1)$ are distinct and $G_2 : \mathbb{Z}_M \to \mathbb{Z}_N$ is a truly random function, this case happens with probability $1/N^2$. From Lemma 4, in this case we always have $(B_{i,j} = 1)$ and $(B_{k,\ell} = 1)$.

Case 2: $\mathsf{RH}(X_i^2) = \mathsf{RH}(X_j^2)$ and $\mathsf{RH}(X_i^2) \neq \mathsf{RH}(X_\ell^2)$. This case happens with probability $\frac{1}{N}\left(1 - \frac{1}{N}\right)$. Now again in this case, from Lemma 4, $B_{i,j} = 1$ with certainty, and $B_{k,\ell} = 1$ with probability $1/M$. Hence in this case the conditional probability that $B_{i,j} = 1$ and $B_{k,\ell} = 1$ is $1/M$.

Case 3: $\mathsf{RH}(X_i^2) \neq \mathsf{RH}(X_j^2)$ and $\mathsf{RH}(X_i^2) = \mathsf{RH}(X_\ell^2)$. This is similar to Case 2.

Case 4: $\mathsf{RH}(X_i^2) \neq \mathsf{RH}(X_j^2)$ and $\mathsf{RH}(X_j^2) = \mathsf{RH}(X_\ell^2)$. This case again happens with probability $\frac{1}{N}\left(1 - \frac{1}{N}\right)$. From Lemma 4, we have

$$\mathsf{LH}(C_j) \boxminus \mathsf{LH}(C_\ell) = \mathsf{LH}(X_j) \boxminus \mathsf{LH}(X_\ell) \ . \tag{25}$$

Assume that $B_{i,j} = 1$, which by Lemma 4 happens with probability $1/M$. We now prove that $B_{i,\ell}$ will be 1. Indeed, since $B_{i,j} = 1$,

$$\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j) = \mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j) \ . \tag{26}$$

Performing the operation $\boxplus$ on Equation (25) and Equation (26) side by side, we obtain

$$\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_\ell) = \mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_\ell) \ .$$

Hence in this case, the conditional probability that $B_{i,j} = 1$ and $B_{k,\ell} = 1$ is $1/M$.

Case 5: $\mathsf{RH}(X_i^2), \mathsf{RH}(X_j^2), \mathsf{RH}(X_\ell^2)$ are distinct. This case happens with probability

$$1 - \frac{1}{N^2} - \frac{3}{N}\left(1 - \frac{1}{N}\right) = 1 - \frac{3}{N} + \frac{2}{N^2} \ .$$

In this case, $\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j)$ is

$$(\mathsf{LH}(X_i^2) \boxplus G_3(\mathsf{RH}(X_i^2))) \boxminus (\mathsf{LH}(X_j^2) \boxplus G_3(\mathsf{RH}(X_j^2))) \ ,$$

and a similar formula holds for $\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_\ell)$. As $\mathsf{RH}(X_i^2), \mathsf{RH}(X_j^2), \mathsf{RH}(X_\ell^2)$ are distinct and G_3 is a truly random function, $\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j)$ and $\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_\ell)$ are independently and uniformly distributed in $\mathbb{Z}_M$. Hence in this case the conditional probability that $B_{i,j} = 1$ and $B_{k,\ell} = 1$ is $1/M^2$.

Combining all cases,

$$\begin{aligned}\Pr[(B_{i,j} = 1) \cap (B_{k,\ell} = 1)] &= \frac{1}{N^2} + \frac{3}{MN}\left(1 - \frac{1}{N}\right) + \left(1 - \frac{3}{N} + \frac{3}{N^2}\right)\frac{1}{M^2} \\ &= \frac{(M+N-1)^2}{(MN)^2} + \frac{MN - M - N + 2}{(MN)^2} \ .\end{aligned}$$

Hence

$$\begin{aligned}\mathbf{Cov}(B_{i,j}, B_{k,\ell}) &= \Pr[(B_{i,j} = 1) \cap (B_{k,\ell} = 1)] - \Pr[B_{i,j} = 1] \cdot \Pr[B_{k,\ell} = 1] \\ &= \frac{MN - M - N + 2}{(MN)^2} \ .\end{aligned}$$

JUSTIFYING EQUATION (23). Since $X_i \neq X_j$ and $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$, we must have $\mathsf{LH}(X_i) \neq \mathsf{LH}(X_j)$. Hence

$$\begin{aligned}\mathsf{LH}(X_i^1) &= \mathsf{LH}(X_i) \boxplus G_1(\mathsf{RH}(X_i)) \\ &\neq \mathsf{LH}(X_j) \boxplus G_1(\mathsf{RH}(X_j)) = \mathsf{LH}(X_j^1) \ .\end{aligned}$$

Repeating this argument, we conclude that the left segments of $X_i^1, X_j^1, X_k^1, X_\ell^1$ are distinct. Thus $\mathsf{RH}(X_i^2), \mathsf{RH}(X_j^2), \mathsf{RH}(X_k^2)$, and $\mathsf{RH}(X_\ell^2)$ are independent and uniformly distributed over $\mathbb{Z}_N$, since they are produced from G_2 on distinct inputs. We consider the following cases.

Case 1: $\mathsf{RH}(X_i^2) = \mathsf{RH}(X_j^2)$ and $\mathsf{RH}(X_k^2) = \mathsf{RH}(X_\ell^2)$. This case happens with probability $1/N^2$. In this case, from Lemma 4, we always have $(B_{i,j} = 1)$ and $(B_{k,\ell} = 1)$.

Case 2: $\mathsf{RH}(X_i^2) = \mathsf{RH}(X_j^2)$ and $\mathsf{RH}(X_k^2) \neq \mathsf{RH}(X_\ell^2)$. This case happens with probability $\frac{1}{N}\left(1 - \frac{1}{N}\right)$. In this case, from Lemma 4, $B_{i,j} = 1$ with certainty, and $B_{k,\ell} = 1$ with probability $1/M$. Hence in this case the conditional probability that $(B_{i,j} = 1)$ and $(B_{k,\ell} = 1)$ is $1/M$.

Case 3: $\mathsf{RH}(X_i^2) \neq \mathsf{RH}(X_j^2)$ and $\mathsf{RH}(X_k^2) = \mathsf{RH}(X_\ell^2)$. This is similar to Case 2.

Case 4: $\mathsf{RH}(X_i^2) \neq \mathsf{RH}(X_j^2)$ and $\mathsf{RH}(X_k^2) \neq \mathsf{RH}(X_\ell^2)$, which happens with probability $(1 - 1/N)^2$. Let Bad be the event that

$$\{\mathsf{RH}(X_i^2), \mathsf{RH}(X_j^2)\} = \{\mathsf{RH}(X_k^2), \mathsf{RH}(X_\ell^2)\} \ ,$$

which happens in this case with conditional probability $\frac{2}{N(N-1)}$.

$\triangleright$ First suppose that Bad does not happen. Then $\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j)$ is

$$(\mathsf{LH}(X_i^2) \boxplus G_3(\mathsf{RH}(X_i^2)) \boxminus (\mathsf{LH}(X_j^2) \boxplus G_3(\mathsf{RH}(X_j^2))$$

and a similar formula holds for $\mathsf{LH}(C_k) \boxminus \mathsf{LH}(C_\ell)$.

Case 4.1: $\{\mathsf{RH}(X_i^2), \mathsf{RH}(X_j^2)\} \cap \{\mathsf{RH}(X_k^2), \mathsf{RH}(X_\ell^2)\} = \emptyset$. As the right segments of $X_i^2, X_j^2, X_k^2, X_\ell^2$ are distinct and G_3 is a truly random function, $\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j)$ and $\mathsf{LH}(C_k) \boxminus \mathsf{LH}(C_\ell)$ are independently and uniformly distributed over $\mathbb{Z}_M$. Hence in this sub-case the conditional probability that $B_{i,j} = 1$ and $B_{k,\ell} = 1$ is $1/M^2$.

Case 4.2: $|\{\mathsf{RH}(X_i^2), \mathsf{RH}(X_j^2)\} \cap \{\mathsf{RH}(X_k^2), \mathsf{RH}(X_\ell^2)\}| = 1$. Without loss of generality, suppose that $\mathsf{RH}(X_k^2) \notin \{\mathsf{RH}(X_i^2), \mathsf{RH}(X_j^2)\}$ and $\mathsf{RH}(X_\ell^2) = \mathsf{RH}(X_i^2)$. Again, $\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j)$ and $\mathsf{LH}(C_k) \boxminus \mathsf{LH}(C_\ell)$ are independently and uniformly distributed over $\mathbb{Z}_M$. Hence in this sub-case the conditional probability that $B_{i,j} = 1$ and $B_{k,\ell} = 1$ is $1/M^2$.

Summing up, if Bad does not happen, the conditional probability that $B_{i,j} = 1$ and $B_{k,\ell} = 1$ is $1/M^2$.

$\triangleright$ Next suppose that Bad does happen. In this sub-case, from Lemma 4, $B_{i,j} = 1$ with probability $1/M$, and thus the conditional probability that $B_{i,j} = 1$ and $B_{k,\ell}$ happens is at most $1/M$.

$\triangleright$ Combining the two sub-cases, in Case 4, the conditional probability that $B_{i,j} = 1$ and $B_{k,\ell} = 1$ is at most

$$\begin{aligned}\Pr[\mathsf{Bad}]\frac{1}{M} + (1 - \Pr[\mathsf{Bad}])\frac{1}{M^2} &= \frac{1}{M^2} + \Pr[\mathsf{Bad}] \cdot \frac{M-1}{M^2} \\ &= \frac{1}{M^2} + \frac{2(M-1)}{M^2N(N-1)} \ .\end{aligned}$$

Combining all cases,

$$\Pr[(B_{i,j} = 1) \cap (B_{k,\ell} = 1)] \leq \Big(\frac{M+N-1}{MN}\Big)^2 + \frac{2(M-1)(N-1)}{M^2N^3} \ .$$

Hence

$$\begin{aligned}\mathbf{Cov}(B_{i,j}, B_{k,\ell}) &\leq \Pr[(B_{i,j} = 1) \cap (B_{k,\ell} = 1)] - \Pr[B_{i,j} = 1] \cdot \Pr[B_{k,\ell} = 1] \\ &\leq \frac{2(M-1)(N-1)}{M^2N^3} \ .\end{aligned}$$

Justifying Equation (24). Again $\mathsf{LH}(X_i^1) \neq \mathsf{LH}(X_j^1)$ and $\mathsf{LH}(X_k^1) \neq \mathsf{LH}(X_\ell^1)$. Let $S = \{\mathsf{LH}(X_i^1), \mathsf{LH}(X_j^1)\} \cap \{\mathsf{LH}(X_k^1), \mathsf{LH}(X_\ell^1)\}$. We consider the following cases.

Case 1: $|S| = 0$. Using the same argument as in the justification of Equation (23), in this case $\mathbf{Cov}(B_{i,j}, B_{k,\ell}) \leq \frac{2(M-1)(N-1)}{M^2N^3}$.

Case 2: $|S| = 1$. Due to symmetry, without loss of generality, assume that $\mathsf{LH}(X_i^1) = \mathsf{LH}(X_k^1)$ but $\mathsf{LH}(X_j^1) \neq \mathsf{LH}(X_\ell^1)$. Then (1) $\mathsf{RH}(X_i^2) \neq \mathsf{RH}(X_k^2)$, (2) $\mathsf{RH}(X_i^2), \mathsf{RH}(X_j^2)$, and $\mathsf{RH}(X_\ell^2)$ are independent and uniformly distributed over $\mathbb{Z}_N$, and (3) $\mathsf{RH}(X_j^2), \mathsf{RH}(X_k^2)$, and $\mathsf{RH}(X_\ell^2)$ are independent and uniformly distributed over $\mathbb{Z}_N$. Using a similar analysis as in the justification of Equation (23), with the observation that now the (conditional) probability of the event Bad becomes just $\frac{1}{N(N-1)}$, in this case, $\mathbf{Cov}(B_{i,j}, B_{k,\ell}) \leq \frac{(M-1)(N-1)}{M^2N^3}$.

Case 3: $|S| = 2$. We consider the following sub-cases.

Case 3.1: $\mathsf{LH}(X_i^1) = \mathsf{LH}(X_k^1)$ and $\mathsf{LH}(X_j^1) = \mathsf{LH}(X_\ell^1)$. Note that

$$\begin{aligned}\mathsf{LH}(X_i^1) - \mathsf{LH}(X_j^1) &= \mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j) \ , \text{ and} \\ \mathsf{LH}(X_k^1) - \mathsf{LH}(X_\ell^1) &= \mathsf{LH}(X_k) \boxminus \mathsf{LH}(X_\ell) \ .\end{aligned}$$

Thus this case happens if and only if $\mathsf{LH}(X_i^1) = \mathsf{LH}(X_k^1)$ and $\mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j) = \mathsf{LH}(X_k) \boxminus \mathsf{LH}(X_\ell)$. Since $\mathsf{LH}(X_k) \boxminus \mathsf{LH}(X_\ell)$ is uniformly distributed over $\mathbb{Z}_M \backslash \{0\}$

independent of X_i, X_j, X_k and their intermediate outputs, this case happens with probability $\frac{1}{M(M-1)}$. In this case, $\mathbf{Cov}(B_{i,j}, B_{k,\ell}) \leq \Pr[B_{i,j} = 1] = \frac{M+N-1}{MN}$.

Case 3.2: $\mathsf{LH}(X_i^1) = \mathsf{LH}(X_\ell^1)$ and $\mathsf{LH}(X_j^1) = \mathsf{LH}(X_k^1)$. This case is similar to Case 3.1.

Summing up, totally,

$$\begin{aligned}\mathbf{Cov}(B_{i,j}, B_{k,\ell}) &\leq \frac{2(M-1)(N-1)}{M^2N^3} + \frac{2}{M(M-1)} \cdot \frac{M+N-1}{MN} \\ &\leq \frac{6.2 \cdot (M-1)(N-1)}{M^2N^3} \ ,\end{aligned}$$

where the last inequality is due to the fact that $M \geq N \geq 64$.

C.7 Proof of Lemma 8

Let (G_1, G_2, G_3, G_4) be the functions specified by the key of the Feistel network. For each message X_r, let X_r^t denote its round-t intermediate value. For part (a), since $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$ and $X_i \neq X_j$, from Lemma 7, X_i and X_j cannot collide at round 1. Likewise, X_k and X_ℓ do not collide at round 1.

For part (b), assume to the contrary that X_j and X_k collide at round 1, meaning that $\mathsf{LH}(X_j^1) = \mathsf{LH}(X_k^1)$. Then on the one hand,

$$\begin{aligned}\mathsf{RH}(X_j^2) \boxminus \mathsf{RH}(X_k^2) &= \Big(G_2(\mathsf{LH}(X_j^1)) \boxplus \mathsf{RH}(X_j)\Big) \boxminus \Big(G_2(\mathsf{LH}(X_k^1)) \boxplus \mathsf{RH}(X_k)\Big) \\ &= \mathsf{RH}(X_j) \boxminus \mathsf{RH}(X_k) \ .\end{aligned}$$

On the other hand, since $\mathsf{LH}(C_j) = \mathsf{LH}(C_k)$,

$$\begin{aligned}\mathsf{RH}(X_j^2) \boxminus \mathsf{RH}(X_k^2) &= \Big(\mathsf{RH}(C_j) \boxminus G_4(\mathsf{LH}(C_j))\Big) \boxminus \Big(\mathsf{RH}(C_k) \boxminus G_4(\mathsf{LH}(C_k))\Big) \\ &= \mathsf{RH}(C_j) \boxminus \mathsf{RH}(C_k) \ .\end{aligned}$$

Hence

$$\mathsf{RH}(X_j) \boxminus \mathsf{RH}(X_k) = \mathsf{RH}(C_j) \boxminus \mathsf{RH}(C_k) \ ,$$

violating the non-degeneracy conditions of the edge $(v_{i,j}, v_{k,\ell})$. Thus X_j and X_k cannot collide at round 1.

For part (c), assume to the contrary that there are two inter-group collisions at round 1. From parts (a) and (b), the collisions must be between X_i and X_k, and between X_j and X_ℓ, meaning that $\mathsf{LH}(X_i^1) = \mathsf{LH}(X_k^1)$ and $\mathsf{LH}(X_j^1) = \mathsf{LH}(X_\ell^1)$. Hence

$$\begin{aligned}\mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j) &= \Big(\mathsf{LH}(X_i^1) \boxminus G_1(\mathsf{RH}(X_i))\Big) \boxminus \Big(\mathsf{LH}(X_j^1) \boxminus G_1(\mathsf{RH}(X_j))\Big) \\ &= \mathsf{LH}(X_i^1) \boxminus \mathsf{LH}(X_j^1) \ ,\end{aligned}$$

and likewise,

$$\mathsf{LH}(X_k) \boxminus \mathsf{LH}(X_\ell) = \mathsf{LH}(X_k^1) \boxminus \mathsf{LH}(X_\ell^1) \ .$$

Since $\mathsf{LH}(X_i^1) = \mathsf{LH}(X_k^1)$ and $\mathsf{LH}(X_j^1) = \mathsf{LH}(X_\ell^1)$,

$$\mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j) = \mathsf{LH}(X_k) \boxminus \mathsf{LH}(X_\ell) \ ,$$

violating the non-degeneracy requirements of the edge $(v_{i,j}, v_{k,\ell})$.

For part (d), assume to the contrary that there is some collision, say between X_i and X_ℓ, at round 2. We claim that

$$G_3(\mathsf{RH}(X_i^2)) = G_3(\mathsf{RH}(X_j^2)), \tag{27}$$

and likewise

$$G_3(\mathsf{RH}(X_k^2)) = G_3(\mathsf{RH}(X_\ell^2)) \ . \tag{28}$$

To justify Equation (27), note that on the one hand, since $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$,

$$\begin{aligned}\mathsf{LH}(X_i^1) \boxminus \mathsf{LH}(X_j^1) &= \Big(G_1(\mathsf{RH}(X_i)) \boxplus \mathsf{LH}(X_i)\Big) \boxminus \Big(G_1(\mathsf{RH}(X_j)) \boxplus \mathsf{LH}(X_j)\Big) \\ &= \mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j) = \mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j) \ ,\end{aligned}$$

where the last equality is due to the fact that $v_{i,j}$ is a node in $\mathcal{G}$. On the other hand,

$$\begin{aligned}\mathsf{LH}(X_i^1) \boxminus \mathsf{LH}(X_j^1) &= \Big(\mathsf{LH}(C_i) \boxminus G_3(\mathsf{RH}(X_i^2))\Big) \boxminus \Big(\mathsf{LH}(C_j) \boxminus G_3(\mathsf{RH}(X_j^2))\Big) \\ &= \Big(\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j)\Big) \boxminus \Big(G_3(\mathsf{RH}(X_i^2)) \boxminus G_3(\mathsf{RH}(X_j^2))\Big) \ .\end{aligned}$$

Hence $G_3(\mathsf{RH}(X_i^2)) \boxminus G_3(\mathsf{RH}(X_j^2))$ must be 0, justifying Equation (27). Now, from Equation (27) and Equation (28), if X_i and X_ℓ collide at round 2, meaning that $\mathsf{RH}(X_i^2) = \mathsf{RH}(X_\ell^2)$, it follows that

$$G_3(\mathsf{RH}(X_j^2)) = G_3(\mathsf{RH}(X_k^2)) \ .$$

Moreover, since $\mathsf{LH}(C_j) = \mathsf{LH}(C_k)$,

$$\mathsf{LH}(X_j^1) = \mathsf{LH}(C_j) \boxminus G_3(\mathsf{RH}(X_j^2)) = \mathsf{LH}(C_k) \boxminus G_3(\mathsf{RH}(X_k^2)) = \mathsf{LH}(X_k^1) \ .$$

In other words, X_j and X_k collide at round 1, contradicting part (b).

C.8 Proof of Lemma 10

For part (a), let $i, j \in \{1, \ldots, p\}$ be arbitrary distinct indices, and let $B_{i,j}$ be the Bernoulli random variable such that $B_{i,j} = 1$ if and only if $\mathcal{G}$ contains node $v_{i,j}$. We now compute $\Pr[B_{i,j} = 1]$. First, since X_i and X_j are sampled uniformly without replacement from $\mathbb{Z}_M \times \mathbb{Z}_N$, the chance that $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$ is $\frac{M-1}{MN-1}$.

Next, given that $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$, from Lemma 3, the conditional probability that $\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j) = \mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j)$ is $\frac{M+N-1}{MN}$. Thus

$$\Pr[B_{i,j} = 1] = \frac{M-1}{MN-1} \cdot \frac{M+N-1}{MN} = \frac{(M-1)(M+N-1)}{MN(MN-1)}$$

Hence

$$\begin{aligned}\mathbf{E}[|V|] = \mathbf{E}\Big[\sum_{i \neq j} B_{i,j}\Big] = \sum_{i \neq j} \mathbf{E}[B_{i,j}] &= \sum_{i \neq j} \Pr[B_{i,j} = 1] \\ &= \frac{p(p-1)(M-1)(M+N-1)}{MN(MN-1)} \ .\end{aligned}$$

For part (b), let $i, j \in \{1, \ldots, p\}$ be arbitrary distinct indices, and let $D_{i,j}$ be the Bernoulli random variable such that $D_{i,j} = 1$ if and only if $\mathcal{G}$ contains node $v_{i,j}$, and this node is good. We now compute $\Pr[D_{i,j} = 1]$. For each $t \in \{1, 2, 3, 4\}$, let X_i^t and X_j^t denote the round-t intermediate outputs of X_i and X_i respectively. Then, from part (b) of Lemma 4, $D_{i,j} = 1$ if and only if $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$ and $\mathsf{RH}(X_i^2) = \mathsf{RH}(X_j^2)$. Again, the chance that $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$ is $\frac{M-1}{MN-1}$. From part (a) of Lemma 4, given that $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$, the conditional probability that $X_i^2 = X_j^2$ is exactly $1/N$. Thus $\Pr[D_{i,j} = 1] = \frac{M-1}{(MN-1)N}$. Hence

$$\begin{aligned}\mathbf{E}[Z] = \mathbf{E}\Big[\sum_{i \neq j} D_{i,j}\Big] = \sum_{i \neq j} \mathbf{E}[D_{i,j}] &= \sum_{i \neq j} \Pr[D_{i,j} = 1] \\ &= \frac{p(p-1)(M-1)}{(MN-1)N} \ .\end{aligned}$$

For part (c), for distinct $i, j, k, \ell \in \{1, \ldots, p\}$, let $B_{i,j,k,\ell}$ denote the Bernoulli random variable such that $B_{i,j,k,\ell} = 1$ if and only if $(v_{i,j}, v_{k,\ell})$ is an edge in $\mathcal{G}$. Then

$$\mathbf{E}[|E|] = \mathbf{E}\Big[\sum_{i,j,k,\ell} B_{i,j,k,\ell}\Big] = \sum_{i,j,k,\ell} \mathbf{E}[B_{i,j,k,\ell}] = \sum_{i,j,k,\ell} \Pr[B_{i,j,k,\ell} = 1] \ .$$

We claim that each $\Pr[B_{i,j,k,\ell} = 1]$ is at most $\frac{(M+N)^2}{M^3N^4}$. Hence

$$\mathbf{E}[|E|] \leq \frac{p!}{(p-4)!} \cdot \frac{(M+N)^2}{M^3N^4} \ .$$

We now justify the claim above. Fix $(i, j, k, \ell) \in (\{1, \ldots, p\})^4$ such that i, j, k, ℓ are distinct. For $B_{i,j,k,\ell} = 1$, the messages X_i, X_j, X_k, X_ℓ have to satisfy the following constraints:

(i) $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$ and $\mathsf{RH}(X_k) = \mathsf{RH}(X_\ell)$, but $\mathsf{RH}(X_j) \neq \mathsf{RH}(X_k)$.
(ii) $\Big(\mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j)\Big) \neq \Big(\mathsf{LH}(X_k) \boxminus \mathsf{LH}(X_\ell)\Big)$.
(iii) $\mathsf{LH}(X_i) \neq \mathsf{LH}(X_j)$ and $\mathsf{LH}(X_k) \neq \mathsf{LH}(X_\ell)$.

First, there are $N(N-1)$ choices for $\mathsf{RH}(X_i), \mathsf{RH}(X_j), \mathsf{RH}(X_k), \mathsf{RH}(X_\ell)$ meeting condition (i). Moreover, there are $M^2(M-1)(M-2)$ choices of $\mathsf{LH}(X_i), \mathsf{LH}(X_j)$, $\mathsf{LH}(X_k), \mathsf{LH}(X_\ell)$ meeting (ii) and (iii). Hence if X_i, X_j, X_k, X_ℓ are sampled uniformly without replacement from $\mathbb{Z}_M \times \mathbb{Z}_N$, the chance that they satisfy the constraints above is

$$\frac{N(N-1)M^2(M-1)(M-2)}{MN(MN-1)(MN-2)(MN-3)} \leq \frac{1}{N^2} \ .$$

Suppose that X_i, X_j, X_k, X_ℓ satisfy the constraints (i), (ii), (iii) above. What remains is to prove that the conditional probability that $B_{i,j,k,\ell} = 1$ is at most $\frac{(M+N)^2}{M^3N^2}$. Let (G_1, G_2, G_3, G_4) be the functions specified by the key of the Feistel network. For each message X_r, let X_r^t denote its round-t intermediate value. From Lemma 8, there is at most one collision at round 1, between either X_i and X_k, or between X_i and X_ℓ, or between X_j and X_ℓ. Without loss of generality, assume that X_i and X_ℓ do not collide at round 1, and X_j and X_ℓ also do not collide at round 1. From part (d) of Lemma 8, there are at most two collisions at round 2. We consider the following cases.

Case 1: There are exactly two collisions at round 2. From Lemma 8, X_i and X_j collide at round 2, and X_k and X_ℓ also collide at round 2. Since $\mathsf{LH}(X_i^1), \mathsf{LH}(X_j^1)$, $\mathsf{LH}(X_\ell^1)$ are distinct, and $\mathsf{LH}(X_k^1)$ and $\mathsf{LH}(X_\ell^1)$ are distinct, and G_2 is a truly random function independent of the round-1 outputs, this case happens with probability at most $1/N^2$. Moreover, since $\mathsf{LH}(C_k) = G_3(\mathsf{RH}(X_k^2)) \boxplus \mathsf{LH}(X_k^2)$ and $\mathsf{LH}(C_j) = G_3(\mathsf{RH}(X_j^2)) \boxplus \mathsf{LH}(X_j^2)$, and $\mathsf{RH}(X_k^2) \neq \mathsf{RH}(X_j^2)$, the conditional probability that $\mathsf{LH}(C_k) = \mathsf{LH}(C_j)$ in this case is $1/M$. Hence the conditional probability that $B_{i,j,k,\ell} = 1$ in this case is at most $1/MN^2$.

Case 2: There is exactly one collisions at round 2. From Lemma 8, this is either the collision between X_i and X_j, or between X_k and X_ℓ. Without loss of generality, suppose that X_i and X_j collide at round 2. Since $\mathsf{LH}(X_i^1)$ and $\mathsf{LH}(X_j^1)$ are distinct, and G_2 is a truly random function independent of the round-1 outputs, this case happens with probability at most $1/N$. Moreover, recall that

(1) $\mathsf{LH}(C_j) = G_3(\mathsf{RH}(X_j^2)) \boxplus \mathsf{LH}(X_j^2)$ and the similar formulas hold for X_k and X_ℓ respectively,
(2) $\mathsf{RH}(X_j), \mathsf{RH}(X_k), \mathsf{RH}(X_\ell)$ are distinct, and
(3) G_3 is a truly random function independent of the round-2 output.

Hence in this case, $\mathsf{LH}(C_j), \mathsf{LH}(C_k), \mathsf{LH}(C_\ell)$ are (conditionally) independent, truly random strings in $\mathbb{Z}_M$. Thus in this case the chance that $\mathsf{LH}(C_j) = \mathsf{LH}(C_k)$ and $\mathsf{LH}(C_k) \boxminus \mathsf{LH}(C_\ell) = \mathsf{LH}(X_k) \boxminus \mathsf{LH}(X_\ell)$ is at most $1/M^2$. Taking into account the two possibilities for the collision at round 2, the conditional probability that $B_{i,j,k,\ell} = 1$ in this case is at most $2/M^2N$.

Case 3: There is no collision at round 2. Then in this case $\mathsf{LH}(C_i), \mathsf{LH}(C_j)$, $\mathsf{LH}(C_k), \mathsf{LH}(C_\ell)$ are (conditionally) independent, truly random strings in $\mathbb{Z}_M$.

Hence in this case the chance that $\mathsf{LH}(C_j) = \mathsf{LH}(C_k)$ and $\mathsf{LH}(C_k) \boxminus \mathsf{LH}(C_\ell) = \mathsf{LH}(X_k) \boxminus \mathsf{LH}(X_\ell)$ and $\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j) = \mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j)$ is at most $1/M^3$. Thus the conditional probability that $B_{i,j,k,\ell} = 1$ in this case is at most $1/M^3$.

Summing over all three cases, the conditional probability that $B_{i,j,k,\ell} = 1$ is at most

$$\frac{1}{MN^2} + \frac{2}{M^2N} + \frac{1}{M^3} = \frac{1}{M}\Big(\frac{1}{M} + \frac{1}{N}\Big)^2 = \frac{(M+N)^2}{M^3N^2} \ .$$

C.9 Proof of Lemma 12

Setup. Let (G_1, G_2, G_3, G_4) be the functions specified by the key of the Feistel network. For each message X_i, let X_i^t denote its round-t intermediate value. Fix $(i, j, k, \ell, r, s) \in (\{1, \ldots, p\})^6$ such that i, j, k, ℓ, r, s are distinct and

(i) $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$, $\mathsf{RH}(X_k) = \mathsf{RH}(X_\ell)$, and $\mathsf{RH}(X_r) = \mathsf{RH}(X_s)$, and
(ii) $\mathsf{RH}(X_i), \mathsf{RH}(X_k)$ and $\mathsf{RH}(X_r)$ are distinct, and
(iii) $\Big(\mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j)\Big) \boxplus \Big(\mathsf{LH}(X_k) \boxminus \mathsf{LH}(X_j)\Big) \boxplus \Big(\mathsf{LH}(X_r) \boxminus \mathsf{LH}(X_s)\Big) = 0.$

The conditions above are necessary so that there exist $(C_i, \ldots, C_s)$ such that $\mathcal{T} = (v_{i,j}, v_{k,\ell}, v_{r,s})$ is a triangle in $\mathcal{G}$.[12] Let Good be the event that $\mathcal{T} = (v_{i,j}, v_{k,\ell}, v_{r,s})$ forms a triangle of zero weight in $\mathcal{G}$, and all the three nodes are good. Let Bad be the event that $\mathcal{T}$ is a triangle of zero weight in $\mathcal{G}$, but some of its nodes is bad. We will prove that

$$\Pr[\mathsf{Bad}] \le \Pr[\mathsf{Good}] \cdot \frac{N}{M-9}\Big(\frac{4}{M} + \frac{33N}{(N-2)M^2} + \frac{39}{M^2}\Big) \ , \text{ and} \tag{29}$$

$$\Pr[\mathsf{Good}] \ge \frac{1}{N^3M^2}\Big(1 - \frac{4}{N} - \frac{9}{M}\Big) \ . \tag{30}$$

Those claims will be justified later below.

For part (a), since Good and Bad are disjoint,

$$\Pr[\mathsf{Good} \mid \mathsf{Good} \cup \mathsf{Bad}] = \frac{\Pr[\mathsf{Good}]}{\Pr[\mathsf{Good} \cup \mathsf{Bad}]} = \frac{\Pr[\mathsf{Good}]}{\Pr[\mathsf{Good}] + \Pr[\mathsf{Bad}]}$$
$$= \frac{1}{1 + \Pr[\mathsf{Bad}]/\Pr[\mathsf{Good}]} \ .$$

Thus from Equation (29),

$$\Pr[\mathsf{Good} \mid \mathsf{Good} \cup \mathsf{Bad}] \ge \frac{1}{1+\epsilon} \ ,$$

[12] To justify condition (iii), note that $\mathcal{T} = (v_{i,j}, v_{k,\ell}, v_{r,s})$ to be a triangle, we must have $\mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j) = \mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j)$, $\mathsf{LH}(X_k) \boxminus \mathsf{LH}(X_\ell) = \mathsf{LH}(C_k) \boxminus \mathsf{LH}(C_\ell)$, and $\mathsf{LH}(X_r) \boxminus \mathsf{LH}(X_s) = \mathsf{LH}(C_r) \boxminus \mathsf{LH}(C_s)$. Summing those formulas side-by-side, and taking into account that $\mathsf{LH}(C_j) = \mathsf{LH}(C_k), \mathsf{LH}(C_\ell) = \mathsf{LH}(C_r), \mathsf{LH}(C_s) = \mathsf{LH}(C_i)$, we obtain (iii).

where $\epsilon = \frac{N}{M-9} \cdot \left(\frac{4}{M} + \frac{33N}{(N-2)M^2} + \frac{39}{M^2}\right)$.

For part (b), for each tuple $\mathcal{L} = (i, j, k, \ell, r, s) \in (\{1, \ldots, p\})^6$ such that i, j, k, ℓ, r, and s are distinct, let $B_{\mathcal{L}}$ be indicator random variable for the event that $\mathcal{T} = (v_{i,j}, v_{k,\ell}, v_{r,s})$ forms a triangle of zero weight in $\mathcal{G}$, and all the three nodes are good. Let Λ denote the number of triangles of zero-weight in $\mathcal{G}$ whose all nodes are good. Then

$$\mathbf{E}[\Lambda] = \frac{1}{3}\mathbf{E}\Big[\sum_{\mathcal{L}} B_{\mathcal{L}}\Big] = \frac{1}{3}\sum_{\mathcal{L}} \mathbf{E}[B_{\mathcal{L}}] = \sum_{\mathcal{L}} \Pr[B_{\mathcal{L}} = 1] \ ,$$

where the factor $1/3$ is due to the fact that each triangle is counted three times from its three nodes in the sum $\sum_{\mathcal{L}} B_{\mathcal{L}}$. For each $\mathcal{L}$, $B_{\mathcal{L}} = 1$ only if (1) $(X_i, \ldots, X_s)$ satisfy conditions (i), (ii), (iii) above, and (2) the event $\mathsf{Good}_{\mathcal{L}}$ happens. To estimate the probability of event (1), first note that there are at least

$$N(N-1)(N-2) \geq N^3\Big(1 - \frac{3}{N}\Big)$$

choices of for $(\mathsf{RH}(X_i), \ldots, \mathsf{RH}(X_s))$ such that $\mathsf{RH}(X_i) = \mathsf{RH}(X_j), \mathsf{RH}(X_k) = \mathsf{RH}(X_\ell), \mathsf{RH}(X_r) = \mathsf{RH}(X_s)$, and $\mathsf{RH}(X_i), \mathsf{RH}(X_k)$, and $\mathsf{RH}(X_r)$ are distinct. Next, there are at least

$$M^3(M-1)(M-4) \geq M^5 \cdot \Big(1 - \frac{5}{M}\Big)$$

choices for $\mathsf{LH}(X_i), \ldots, \mathsf{LH}(X_s)$ such that

- $\mathsf{LH}(X_i) \neq \mathsf{LH}(X_j)$, $\mathsf{LH}(X_k) \neq \mathsf{LH}(X_\ell)$, $\mathsf{LH}(X_r) \neq \mathsf{LH}(X_s)$, and
- $\mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j), \mathsf{LH}(X_k) \boxminus \mathsf{LH}(X_\ell), \mathsf{LH}(X_r) \boxminus \mathsf{LH}(X_s)$ are distinct, and
- $\Big(\mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j)\Big) \boxplus \Big(\mathsf{LH}(X_k) \boxminus \mathsf{LH}(X_j)\Big) \boxplus \Big(\mathsf{LH}(X_r) \boxminus \mathsf{LH}(X_s)\Big) = 0.$

To see why, let $\Delta_1 = \mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j), \Delta_2 = \mathsf{LH}(X_k) \boxminus \mathsf{LH}(X_\ell)$, and $\Delta_3 = \mathsf{LH}(X_r) \boxminus \mathsf{LH}(X_s)$. The constraints above mean that

(i) $\Delta_1, \Delta_2, \Delta_3$ are distinct and non-zero, and
(ii) $\Delta_1 \boxplus \Delta_2 \boxplus \Delta_3 = 0$.

Since there are M^3 choices of the triple $\Big(\mathsf{LH}(X_i), \mathsf{LH}(X_k), \mathsf{LH}(X_r)\Big)$, and $\mathsf{LH}(X_j)$, $\mathsf{LH}(X_\ell), \mathsf{LH}(X_s)$ are completely determined from $\mathsf{LH}(X_i), \mathsf{LH}(X_k), \mathsf{LH}(X_r)$, Δ_1, Δ_2, Δ_3, it suffices to show that there are at least $(M-1)(M-4)$ triples $(\Delta_1, \Delta_2, \Delta_3)$ meeting the constraints (i) and (ii) above. For any element $X \in \mathbb{Z}_M$, let $-X$ denote $0 \boxminus X$ and $-2X$ denote $0 \boxminus (X \boxplus X)$. By substituting $\Delta_3 = -(\Delta_1 \boxplus \Delta_2)$ as specified by (ii), the condition (i) becomes (iii) $\Delta_1 \neq \{0, -2\Delta_2\}$ and $\Delta_2 \notin \{0, -\Delta_1, -2\Delta_1\}$. Since there are at least $(M-1)(M-3)$ pairs $(\Delta_1, \Delta_2) \in (\mathbb{Z}_M \backslash \{0\})^2$ such that $\Delta_2 \notin \{-\Delta_1, -2\Delta_1\}$, and there are exactly $M-1$ pairs $(\Delta_1, \Delta_2) \in (\mathbb{Z}_M \backslash \{0\})^2$ such that $\Delta_1 = -2\Delta_2$, there are at least

$$(M-1)(M-3) - (M-1) = (M-1)(M-4)$$

pairs $(\Delta_1, \Delta_2) \in (\mathbb{Z}_M)^2$ meeting the constraint (iii).

Hence the chance that event (1) happens is at least

$$\frac{M^5N^3(1-3/N)(1-5/M)}{MN(MN-1)\cdots(MN-5)} \geq \frac{1}{MN^3}\Big(1-\frac{3}{N}-\frac{5}{M}\Big) \ .$$

Next, from Equation (30), the conditional probability that event (2) happens, given event (1), is at least $\frac{1}{N^3M^2}\Big(1-\frac{4}{N}-\frac{9}{M}\Big)$. Hence

$$\Pr[B_{\mathcal{L}}=1] \geq \frac{1}{M^3N^6}\Big(1-\frac{7}{N}-\frac{14}{M}\Big) \ ,$$

and thus

$$\mathbf{E}[\Lambda] \geq \frac{p!}{3(p-6)!}\cdot\frac{1}{M^3N^6}\Big(1-\frac{7}{N}-\frac{14}{M}\Big)$$

justifying part (b).

The number of bad nodes in a triangle. We first prove by contradiction that if Bad happens then at least two nodes of the triangle $\mathcal{T}$ are bad. Assume to the contrary that Bad happens but exactly one node of $\mathcal{T}$ is bad. Without loss of generality, assume that this bad node is $v_{r,s}$. Since both $v_{i,j}$ and $v_{k,\ell}$ are good, from Lemma 9,

$$G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_j)) = \mathsf{Label}(v_{i,j})$$
$$G_4(\mathsf{LH}(C_k) \boxminus G_4(\mathsf{LH}(C_\ell)) = \mathsf{Label}(v_{k,\ell}) \ .$$

Adding the two equations above side by side, and observing that $\mathsf{LH}(C_j) = \mathsf{LH}(C_k)$, we obtain

$$G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_\ell)) = \mathsf{Label}(v_{i,j}) \boxplus \mathsf{Label}(v_{k,\ell}) \ .$$

On the other hand, since $\mathsf{LH}(C_\ell) = \mathsf{LH}(C_r)$ and $\mathsf{LH}(C_s) = \mathsf{LH}(C_i)$,

$$G_4(\mathsf{LH}(C_s)) \boxminus G_4(\mathsf{LH}(C_r)) = \mathsf{Label}(v_{i,j}) \boxplus \mathsf{Label}(v_{k,\ell}) \ .$$

Since $\mathcal{T}$ has zero weight, $\mathsf{Label}(v_{i,j}) \boxplus \mathsf{Label}(v_{k,\ell}) = 0 \boxminus \mathsf{Label}(v_{r,s})$. Hence

$$G_4(\mathsf{LH}(C_r)) \boxminus G_4(\mathsf{LH}(C_s)) = \mathsf{Label}(v_{r,s}) \ .$$

Since $G_4(\mathsf{LH}(C_r)) = \mathsf{RH}(C_r) \boxminus \mathsf{RH}(X_r^2)$ and $G_4(\mathsf{LH}(C_s)) = \mathsf{RH}(C_s) \boxminus \mathsf{RH}(X_s^2)$,

$$\Big(\mathsf{RH}(C_r) \boxminus \mathsf{RH}(X_r^2)\Big) \boxminus \Big(\mathsf{RH}(C_s) \boxminus \mathsf{RH}(X_s^2)\Big) = \mathsf{Label}(v_{r,s}) \ . \tag{31}$$

By expanding $\mathsf{Label}(v_{r,s}) = \mathsf{RH}(C_r) \boxminus \mathsf{RH}(C_s)$, Equation (31) above can be simplified as

$$\mathsf{RH}(X_r^2) = \mathsf{RH}(X_s^2) \ .$$

In other words, $v_{r,s}$ is also good, which is a contradiction.

Characterizing collisions. We will now characterize the collisions among the queries.

$\triangleright$ First, from Lemma 8, X_i and X_j cannot collide at round 1, and neither do X_k and X_ℓ, and neither do X_r and X_s. Next, from Lemma 8, X_j and X_k do not collide at round 1, and neither do X_ℓ and X_r, and neither do X_s and X_i.

$\triangleright$ Next, suppose that Bad happens. From Lemma 8,

(a) If all nodes of $\mathcal{T}$ are bad then there is no collision at round 2.
(b) If exactly one node of $\mathcal{T}$, say $v_{i,j}$, is good, then the only collision at round 2 is between X_i and X_j.

$\triangleright$ Next partition the six queries into three groups: $\{X_i, X_j\}, \{X_k, X_\ell\}, \{X_r, X_s\}$. Thus at round 1, there is no intra-group collisions. From Lemma 8, between any two groups, there is at most one inter-group collision at round 1.

$\triangleright$ Suppose that $S = \{\mathsf{LH}(X_i^1), \mathsf{LH}(X_j^1), \mathsf{LH}(X_k^1), \mathsf{LH}(X_\ell^1), \mathsf{LH}(X_r^1), \mathsf{LH}(X_s^1)\}$ contains exactly three elements. We now prove that among $\mathsf{LH}(X_i^1), \mathsf{LH}(X_j^1), \mathsf{LH}(X_k^1), \mathsf{LH}(X_\ell^1), \mathsf{LH}(X_r^1), \mathsf{LH}(X_s^1)$, each value repeats at most twice. Assume to the contrary that this claim is not true. Then there are only two possible cases.

Case 1: Among $\mathsf{LH}(X_i^1), \mathsf{LH}(X_j^1), \mathsf{LH}(X_k^1), \mathsf{LH}(X_\ell^1), \mathsf{LH}(X_r^1), \mathsf{LH}(X_s^1)$, there is some value that repeats exactly three times. Due to the first characterization, this value is either $\mathsf{LH}(X_i^1), \mathsf{LH}(X_k^1), \mathsf{LH}(X_r^1)$, or $\mathsf{LH}(X_j^1), \mathsf{LH}(X_\ell^1), \mathsf{LH}(X_s^1)$. Without loss of generality, suppose that $\mathsf{LH}(X_i^1), \mathsf{LH}(X_k^1), \mathsf{LH}(X_r^1)$ are the same. On the one hand, since $|S| = 3$, there are at least three collisions at round 1. On the other hand, from the first and third characterizations, at round 1, there are at most three collisions. Hence there are exactly three collisions at round 1. This means that $\mathsf{LH}(X_j^1), \mathsf{LH}(X_\ell^1), \mathsf{LH}(X_s^1)$ are distinct, and they are all different from $\mathsf{LH}(X_i^1)$. Hence the set S contains at least 4 elements, which is a contradiction.

Case 2: Among $\mathsf{LH}(X_i^1), \mathsf{LH}(X_j^1), \mathsf{LH}(X_k^1), \mathsf{LH}(X_\ell^1), \mathsf{LH}(X_r^1), \mathsf{LH}(X_s^1)$, there is some value that repeats four times or more. By Pigeonhole principle, among the three groups $\{X_i, X_j\}, \{X_k, X_\ell\}, \{X_r, X_s\}$, there must be two groups such that between them there are two collisions, violating the third characterization.

$\triangleright$ Finally partition the six queries into three groups $\{X_s, X_i\}, \{X_j, X_k\}, \{X_\ell, X_r\}$. We now show that at round 1, between any two groups, there cannot be two inter-group collisions that involve the same query. Assume to the contrary that there are two inter-group collisions between, say the first and second groups, that involve the same queries. By symmetry, without loss of generality, assume that the collisions are between X_s and X_j, and between X_s and X_k. On the one hand, since $\mathsf{LH}(X_s^1) = \mathsf{LH}(X_j^1)$,

$$\begin{aligned}\mathsf{RH}(X_s^2) \boxminus \mathsf{RH}(X_j^2) &= \Big(G_2(\mathsf{LH}(X_s^1) \boxplus \mathsf{RH}(X_s)\Big) \boxminus \Big(G_2(\mathsf{LH}(X_j^1) \boxplus \mathsf{RH}(X_j)\Big) \\ &= \mathsf{RH}(X_s) \boxminus \mathsf{RH}(X_j) \ .\end{aligned}$$

On the other hand,

$$\mathsf{RH}(X_s^2) \boxminus \mathsf{RH}(X_j^2) = \Big(\mathsf{RH}(C_s) \boxminus G_4(\mathsf{LH}(C_s))\Big) \boxminus \Big(\mathsf{RH}(C_j) \boxminus G_4(\mathsf{LH}(C_j))\Big) \ .$$

Hence

$$G_4(\mathsf{LH}(C_s)) \boxminus G_4(\mathsf{LH}(C_j)) = \Big(\mathsf{RH}(C_s) \boxminus \mathsf{RH}(C_j)\Big) \boxminus \Big(\mathsf{RH}(X_s) \boxminus \mathsf{RH}(X_j)\Big) \ .$$

Likewise,

$$G_4(\mathsf{LH}(C_s)) \boxminus G_4(\mathsf{LH}(C_k)) = \Big(\mathsf{RH}(C_s) \boxminus \mathsf{RH}(C_k)\Big) \boxminus \Big(\mathsf{RH}(X_s) \boxminus \mathsf{RH}(X_k)\Big) \ .$$

Since $\mathsf{LH}(C_j) = \mathsf{LH}(C_k)$,

$$\mathsf{RH}(C_j) \boxminus \mathsf{RH}(X_j) = \mathsf{RH}(C_k) \boxminus \mathsf{RH}(X_k) \ ,$$

violating the non-degeneracy requirements of the edge $(v_{i,j}, v_{k,\ell})$.

Proof ideas. Let $\mathcal{D}$ be the set of all possible $(C_i, \ldots, C_s)$ such that $\mathcal{T}$ is a zero-weight triangle of $\mathcal{G}$. On the one hand, we will show that

$$\Pr[\mathsf{Good}] \geq |\mathcal{D}| \cdot \frac{1}{N^8 M^3}\Big(1 - \frac{9}{M}\Big) \ .$$

Our approach is to fix an arbitrary element in $\mathcal{D}$ for $(C_i, \ldots, C_s)$, and then imagine that at the front, we apply the plaintexts with a 2-round Feistel network, and at the back, we apply the ciphertexts with the inverse of a 2-round Feistel. On the other hand, to bound $\Pr[\mathsf{Bad}]$, we will consider several cases based on the collisions at the first and last rounds. In the d-th case, we will show that Bad happens only for a corresponding subset $\mathcal{D}_d \subseteq \mathcal{D}$. We then fix an arbitrary element of $\mathcal{D}_d$ for $(C_i, \ldots, C_s)$ and also bound the probability that two Feistel networks meet properly at the middle by a certain number ε_d. Thus

$$\Pr[\mathsf{Bad}] \leq |\mathcal{D}| \cdot \sum_d \frac{|\mathcal{D}_d|}{|\mathcal{D}|} \cdot \varepsilon_d \ .$$

Bounding the chance of Good. We first give a lower bound of $\Pr[\mathsf{Good}]$. Fix an arbitrary element in $\mathcal{D}$ for $(C_i, \ldots, C_s)$. First, recall that X_i and X_j are of the same right segment, and so do X_k and X_ℓ, and X_r and X_s. Hence from Lemma 7 and the characterization of collisions, the chance that no two queries collide at round 1 is at least

$$1 - \frac{1}{M} - \frac{2}{M} - \frac{3}{M} - \frac{3}{M} = 1 - \frac{9}{M} \ .$$

To see why, the only possible collisions at round 1 are: (i) between X_k and X_i, which happens with probability at most $1/M$, (ii) between X_ℓ and one element of $\{X_i, X_j\}$, which happens with probability at most $2/M$, (iii) between X_r and one element of $\{X_i, X_j, X_k\}$, which happens with probability at most $3/M$, and (iv) between X_s and one element of $\{X_j, X_k, X_r\}$, which happens with probability at most $3/M$.

Next, assume that we have no collision at round 1. Since $\mathsf{LH}(X_i^1), \mathsf{LH}(X_j^1)$, $\mathsf{LH}(X_k^1), \mathsf{LH}(X_\ell^1), \mathsf{LH}(X_r^1), \mathsf{LH}(X_s^1)$ are distinct and since G_2 is a truly random function, the chance that at the front, the following events happen is $\frac{1}{N^5}$:

(1) X_i and X_j collide at round 2,
(2) X_k and X_ℓ collide at round 2,
(3) X_r and X_s collide at round 2,
(4) $\mathsf{RH}(X_j^2) \boxminus \mathsf{RH}(X_k^2) = \mathsf{RH}(C_j) \boxminus \mathsf{RH}(C_k)$,
(5) $\mathsf{RH}(X_\ell^2) \boxminus \mathsf{RH}(X_r^2) = \mathsf{RH}(C_\ell) \boxminus \mathsf{RH}(C_r)$.

Suppose that the five events above all happen. They then imply that

$$\mathsf{RH}(X_s^2) \boxminus \mathsf{RH}(X_i^2) = \mathsf{RH}(C_s) \boxminus \mathsf{RH}(C_i) \ . \tag{32}$$

Moreover, since $\mathsf{LH}(C_j) \neq \mathsf{LH}(C_k)$, it follows that $\mathsf{RH}(X_j^2) \neq \mathsf{RH}(X_k^2)$. Likewise, $\mathsf{RH}(X_\ell^2) \neq \mathsf{RH}(X_r^2)$ and $\mathsf{RH}(X_s^2) \neq \mathsf{RH}(X_i^2)$. Let $(L_i, R_i), \ldots, (L_s, R_s)$ denote the outputs produced by the two-round Feistel network at the front on inputs $X_i, \ldots, X_s$ respectively; note that $R_i = R_j$, $R_k = R_\ell$, $R_r = R_s$, and R_i, R_k, R_r are distinct. Now we need to bound the probability that the outputs at the back are exactly $(L_i, R_i), \ldots, (L_s, R_s)$. At the back, since $\mathsf{LH}(C_i), \mathsf{LH}(C_k), \mathsf{LH}(C_r)$ are distinct and G_4 is a truly random function, the following events happen with (conditional) probability $1/N^3$:

(a) $\mathsf{RH}(C_i) \boxminus G_4(\mathsf{LH}(C_i)) = R_i$,
(b) $\mathsf{RH}(C_k) \boxminus G_4(\mathsf{LH}(C_k)) = R_k$,
(c) $\mathsf{RH}(C_r) \boxminus G_4(\mathsf{LH}(C_r)) = R_r$.

Suppose that the three events above happen. Since $\mathsf{LH}(C_i) = \mathsf{LH}(C_s)$ and from Equation (32), $R_s \boxminus R_i = \mathsf{RH}(C_s) \boxminus \mathsf{RH}(C_i)$, we obtain

$$\mathsf{RH}(C_s) \boxminus G_4(\mathsf{LH}(C_s)) = R_s \ .$$

Likewise,

$$\mathsf{RH}(C_\ell) \boxminus G_4(\mathsf{LH}(C_\ell)) = R_\ell \ ,$$
$$\mathsf{RH}(C_j) \boxminus G_4(\mathsf{LH}(C_j)) = R_j \ .$$

Thus the right segments of the outputs produced by the back two-round inverse Feistel match what were produced by the front. Now, at the back, since R_i, R_k, R_r are distinct and G_3 is a truly random function, the following events happen with (conditional) probability $1/M^3$:

(i) $\mathsf{LH}(C_i) \boxminus G_3(R_i) = L_i$,
(ii) $\mathsf{LH}(C_k) \boxminus G_3(R_k) = L_k$,
(iii) $\mathsf{LH}(C_r) \boxminus G_3(R_r) = L_r$.

Suppose that the three events above happen. Then since $R_i = R_j$, from event (i),

$$\begin{aligned} \mathsf{LH}(C_j) \boxminus G_3(R_j) &= L_i \boxplus \Big(\mathsf{LH}(C_j) \boxminus \mathsf{LH}(C_i)\Big) \\ &= L_i \boxplus \Big(\mathsf{LH}(X_j) \boxminus \mathsf{LH}(X_i)\Big) = L_j \ , \end{aligned}$$

where the second equality follows from the fact that $v_{i,j} \in V$, and the third equality follows from the fact that (1) $L_i = G_1(\mathsf{RH}(X_i)) \boxplus \mathsf{LH}(X_i)$, (2) $L_j = G_1(\mathsf{RH}(X_j)) \boxplus \mathsf{LH}(X_j)$ and (3) $\mathsf{RH}(X_i) = \mathsf{RH}(X_j)$. Likewise,

$$\mathsf{LH}(C_\ell) \boxminus G_3(R_\ell) = L_\ell \ ,$$
$$\mathsf{LH}(C_s) \boxminus G_3(R_s) = L_s \ .$$

Thus the left segments of the outputs produced by the back two-round inverse Feistel match what were produced by the front. Totally, when we restrict the ciphertexts, Good happens with conditional probability at least $\frac{1}{N^8M^3}\Big(1 - \frac{9}{M}\Big)$. Hence,

$$\Pr[\mathsf{Good}] \geq |\mathcal{D}| \cdot \frac{1}{N^8M^3}\Big(1 - \frac{9}{M}\Big) \ . \tag{33}$$

Now, to justify Equation (30), we only need to give a lower bound for $|\mathcal{D}|$. First, there are at least M choices for $(\mathsf{LH}(C_i), \ldots, \mathsf{LH}(C_s))$ such that

- $\mathsf{LH}(C_i), \mathsf{LH}(C_k), \mathsf{LH}(C_r)$ are distinct, and
- $\mathsf{LH}(C_j) = \mathsf{LH}(C_k), \mathsf{LH}(C_\ell) = \mathsf{LH}(C_r), \mathsf{LH}(C_s) = \mathsf{LH}(C_i)$, and
- $\mathsf{LH}(C_i) \boxminus \mathsf{LH}(C_j) = \mathsf{LH}(X_i) \boxminus \mathsf{LH}(X_j)$, $\mathsf{LH}(C_k) \boxminus \mathsf{LH}(C_\ell) = \mathsf{LH}(X_k) \boxminus \mathsf{LH}(X_\ell)$, and $\mathsf{LH}(C_r) \boxminus \mathsf{LH}(C_s) = \mathsf{LH}(X_r) \boxminus \mathsf{LH}(X_s)$.

Next, there are at least

$$N \cdot N \cdot (N-1)(N-1) \cdot (N-2) \geq N^5\Big(1 - \frac{4}{N}\Big)$$

choices for $(\mathsf{RH}(C_i), \ldots, \mathsf{RH}(C_s))$ such that

- $\mathsf{RH}(C_j) \boxminus \mathsf{RH}(C_k) \neq \mathsf{RH}(X_j) \boxminus \mathsf{RH}(X_k)$, $\mathsf{RH}(C_\ell) \boxminus \mathsf{RH}(C_r) \neq \mathsf{RH}(X_\ell) \boxminus \mathsf{RH}(X_r)$, and $\mathsf{RH}(C_s) \boxminus \mathsf{RH}(C_i) \neq \mathsf{RH}(X_s) \boxminus \mathsf{RH}(X_i)$, and
- $\Big(\mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_j)\Big) \boxplus \Big(\mathsf{RH}(C_k) \boxminus \mathsf{RH}(C_\ell)\Big) \boxplus \Big(\mathsf{RH}(C_r) \boxminus \mathsf{RH}(C_s)\Big) = 0$.

Hence,

$$|\mathcal{D}| \geq MN^5\Big(1 - \frac{4}{N}\Big) \ ,$$

and thus

$$\Pr[\mathsf{Good}] \geq \frac{1}{N^3M^2}\Big(1 - \frac{4}{N} - \frac{9}{M}\Big)$$

justifying Equation (30).

Bounding the chance of Bad. We next give an upper bound of $\Pr[\mathsf{Bad}]$. From the characterizations of collisions, the set $S = \{\mathsf{LH}(X_i^1), \mathsf{LH}(X_j^1), \mathsf{LH}(X_k^1), \mathsf{LH}(X_\ell^1), \mathsf{LH}(X_r^1), \mathsf{LH}(X_s^1)\}$ contains at least three distinct elements, and the set $S^* = \{\mathsf{LH}(X_i^2), \mathsf{LH}(X_j^2), \mathsf{LH}(X_k^2), \mathsf{LH}(X_\ell^2), \mathsf{LH}(X_r^2), \mathsf{LH}(X_s^2)\}$ contains at least 5 elements.

Below, we will consider several cases. In each case, we will fix the values of $(C_i, \ldots, C_s)$ from a certain subset of $\mathcal{D}$. Let $L_i, \ldots, L_s$ be the left segments of

the round-1 intermediate outputs of the two-round Feistel network at the front on inputs $X_i, \ldots, X_s$ respectively. Let $R_i, \ldots, R_s$ be the right segments of the round-1 intermediate outputs of the two-round inverse Feistel at the back on inputs $C_i, \ldots, C_s$ respectively. We then bound the probability that the right segments of the outputs at the front are $R_i, \ldots, R_s$, and the left segments of the outputs at the back are $L_i, \ldots, L_s$.

Case 1: $|S| = 6$ and $|S^*| = 6$. Fix an arbitrary element of $\mathcal{D}$ for $(X_i, C_i, \ldots, X_s, C_s)$. In this case $L_i, \ldots, L_s$ are distinct, and $R_i, \ldots, R_s$ are distinct, and thus the chance that $G_2(L_t) \boxplus \mathsf{RH}(X_t) = R_t$ and $G_3(R_t) \boxplus L_t = \mathsf{LH}(C_t)$ for every $t \in \{i, j, k, \ell, r, s\}$ happen with probability $\frac{1}{M^6 N^6}$. Hence in this case, Bad happens with probability at most $|\mathcal{D}| \cdot \frac{1}{M^6 N^6} \leq |\mathcal{D}| \cdot \frac{1}{M^5 N^7}$.

Case 2: $|S| = 6$ and $|S^*| = 5$. From the characterization of collisions, there is exactly one good node in $\mathcal{T}$, say $v_{i,j}$, and the only collision at round 2 is between X_i and X_j. From Lemma 9, in this case we require that

$$G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_j)) = \mathsf{Label}(v_{i,j}) \ .$$

Fix an arbitrary element of $\mathcal{D}$ for $(C_i, \ldots, C_s)$. In this case $L_i, \ldots, L_s$ are distinct, and $R_j, \ldots, R_s$ are distinct. Hence the chance that (1) $G_2(L_t) \boxplus \mathsf{RH}(X_t) = R_t$ for every $t \in \{i, j, k, \ell, r, s\}$, (2) $G_3(R_m) \boxplus L_m = \mathsf{LH}(C_m)$ for every $m \in \{j, k, \ell, r, s\}$, and (3) $G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_j)) = \mathsf{Label}(v_{i,j})$ is $\frac{1}{M^5 N^7}$. By accounting for three possibilities of what the good node in $\mathcal{T}$ is, in this case Bad happens with probability at most $|\mathcal{D}| \cdot \frac{3}{M^5 N^7}$.

Case 3: $|S| = 5$ and $|S^*| = 6$. Thus there is one collision at round 1, and without loss of generality, assume that this collision is between X_i and X_k. This means that

$$\begin{aligned} \mathsf{LH}(X_i^2) \boxminus \mathsf{LH}(X_k^2) &= \Big(G_2(\mathsf{LH}(X_i^1)) \boxplus \mathsf{RH}(X_i)\Big) \boxminus \Big(G_2(\mathsf{LH}(X_k^1)) \boxplus \mathsf{RH}(X_k)\Big) \\ &= \mathsf{RH}(X_i) \boxminus \mathsf{RH}(X_k) \ . \end{aligned}$$

On the other hand,

$$\mathsf{LH}(X_i^2) \boxminus \mathsf{LH}(X_k^2) = \Big(\mathsf{RH}(C_i) \boxminus G_4(\mathsf{LH}(C_i))\Big) \boxminus \Big(\mathsf{RH}(C_k) \boxminus G_4(\mathsf{LH}(C_k))\Big) \ .$$

Hence

$$G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_k)) = \Big(\mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_k)\Big) \boxminus \Big(\mathsf{RH}(X_i) \boxminus \mathsf{RH}(X_k)\Big) \ .$$

Moreover, since X_i collides with X_k at round 1,

$$G_1(\mathsf{RH}(X_i)) \boxplus \mathsf{LH}(X_i) = G_1(\mathsf{RH}(X_k)) \boxplus \mathsf{LH}(X_k) \ .$$

Fix an arbitrary element of $\mathcal{D}$ for $(C_i, \ldots, C_s)$. In this case $L_j, \ldots, L_k$ are distinct, and $R_i, \ldots, R_k$ are distinct. Hence the chance that the following events happen is $\frac{1}{M^7 N^6}$:

(1) $G_2(L_t) \boxplus \mathsf{RH}(X_t) = R_t$ for every $t \in \{j, k, \ell, r, s\}$,

(2) $G_3(R_m) \boxplus L_m = \mathsf{LH}(C_m)$ for every $m \in \{i, j, k, \ell, r, s\}$,
(3) $G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_k)) = \Big(\mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_k)\Big) \boxminus \Big(\mathsf{RH}(X_i) \boxminus \mathsf{RH}(X_k)\Big)$,
(4) $G_1(\mathsf{RH}(X_i)) \boxplus \mathsf{LH}(X_i) = G_1(\mathsf{RH}(X_k)) \boxplus \mathsf{LH}(X_k)$.

By accounting for the nine possibilities of the collision at round 1, in this case the probability that Bad happens is at most $|\mathcal{D}| \cdot \frac{9}{M^7N^6} \leq |\mathcal{D}| \cdot \frac{9}{M^6N^7}$.

Case 4: $|S| = 5$ and $|S^*| = 5$. From the characterization of collisions, there is exactly one good node in $\mathcal{T}$, say $v_{i,j}$, and the only collision at round 2 is between X_i and X_j. We consider the following sub-cases.

Case 4.1: The collision at round 1 is either between X_i and X_k, or between X_s and X_j, or between X_s and X_k. Suppose that the collision at round 1 is between X_i and X_k; the other cases are similar. This then leads to an equation of G_1 and another equation of G_4:

$$G_1(\mathsf{RH}(X_i)) \boxplus \mathsf{LH}(X_i) = G_1(\mathsf{RH}(X_k)) \boxplus \mathsf{LH}(X_k)$$
$$G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_k)) = \Big(\mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_k)\Big) \boxminus \Big(\mathsf{RH}(X_i) \boxminus \mathsf{RH}(X_k)\Big)$$

On the other hand, the collision between X_i and X_j at round 2 leads to another equation of G_4:

$$G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_j)) = \mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_j) \ .$$

Since $\mathsf{LH}(C_j) = \mathsf{LH}(C_k)$, the two equations above on G_4 imply that

$$\mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_j) = \Big(\mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_k)\Big) \boxminus \Big(\mathsf{RH}(X_i) \boxminus \mathsf{RH}(X_k)\Big) \ . \tag{34}$$

Let $\mathcal{D}_1$ be the subset of $\mathcal{D}$ such that any $(C_i, \ldots, C_s) \in \mathcal{D}_1$ satisfies Equation (34) above. Note that

$$\frac{|\mathcal{D}_1|}{|\mathcal{D}|} \leq \frac{1}{N-2} \ .$$

To see why, fix $C_i, C_j, C_r, C_s, \mathsf{LH}(C_k), \mathsf{LH}(C_\ell)$ such that there exists at least one corresponding $(C_i, \ldots, C_s) \in \mathcal{D}$. Any pair $(\mathsf{RH}(C_k), \mathsf{RH}(C_\ell))$ that makes $(C_i, \ldots, C_s)$ fall in $\mathcal{D}$ is called *compatible*. Then there are at least $N-2$ pairs $(\mathsf{RH}(C_k), \mathsf{RH}(C_\ell)) \in (\mathbb{Z}_N)^2$ such that

$$\Big(\mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_j)\Big) \boxplus \Big(\mathsf{RH}(C_k) \boxminus \mathsf{RH}(C_\ell)\Big) \boxplus \Big(\mathsf{RH}(C_r) \boxminus \mathsf{RH}(C_s)\Big) = 0 \tag{35}$$

and $\mathsf{RH}(C_k) \neq \mathsf{RH}(C_j)$, and $\mathsf{RH}(C_\ell) \neq \mathsf{RH}(C_r)$. and all these pairs are compatible. On the other hand, there is at most one pair $(\mathsf{RH}(C_k), \mathsf{RH}(C_\ell))$ that satisfies both Equation (34) and Equation (35).

Next, fix an arbitrary element of $\mathcal{D}$ for $(C_i, \ldots, C_s)$. In this case $L_j, \ldots, L_k$ are distinct, and $R_j, \ldots, R_k$ are distinct. Hence the chance that the following events happen is $\frac{1}{M^6N^6}$:

(1) $G_2(L_t) \boxplus \mathsf{RH}(X_t) = R_t$ and $G_3(R_t) \boxplus L_t = \mathsf{LH}(C_t)$ for every $t \in \{j, k, \ell, r, s\}$,

(2) $G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_k)) = \Big(\mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_k)\Big) \boxminus \Big(\mathsf{RH}(X_i) \boxminus \mathsf{RH}(X_k)\Big)$,
(3) $G_1(\mathsf{RH}(X_i)) \boxplus \mathsf{LH}(X_i) = G_1(\mathsf{RH}(X_k)) \boxplus \mathsf{LH}(X_k)$.

By accounting for three possibilities of the good node in $\mathcal{T}$, and three corresponding possibilities of the collision at round 1, in Case 4.1, Bad happens with probability at most $|\mathcal{D}_1| \cdot \frac{9}{M^6N^6} \leq |\mathcal{D}| \cdot \frac{9}{(N-2)M^6N^6}$.

Case 4.2: The collision at round 1 is neither between X_i and X_k, nor between X_s and X_j, nor between X_s and X_k. Suppose that the collision at round 1 is between X_i and X_r; the other cases are similar. This then leads to an equation of G_1 and another equation of G_4:

$$G_1(\mathsf{RH}(X_i)) \boxplus \mathsf{LH}(X_i) = G_1(\mathsf{RH}(X_k)) \boxplus \mathsf{LH}(X_k)$$
$$G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_r)) = \Big(\mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_r)\Big) \boxminus \Big(\mathsf{RH}(X_i) \boxminus \mathsf{RH}(X_k)\Big)$$

On the other hand, the collision between X_i and X_j at round 2 leads to another equation of G_4:

$$G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_j)) = \mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_j) \ .$$

Note that the two equations on G_4 are linearly independent. Fix an arbitrary element of $\mathcal{D}$ for $(C_i, \ldots, C_s)$. In this case $L_j, \ldots, L_k$ are distinct, and $R_j, \ldots, R_k$ are distinct. Hence the chance that the following events happen is $\frac{1}{M^6N^7}$:

(1) $G_2(L_t) \boxplus \mathsf{RH}(X_t) = R_t$ and $G_3(R_t) \boxplus L_t = \mathsf{LH}(C_t)$ for every $t \in \{j, k, \ell, r, s\}$,
(2) $G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_r)) = \Big(\mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_r)\Big) \boxminus \Big(\mathsf{RH}(X_i) \boxminus \mathsf{RH}(X_k)\Big)$,
(4) $G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_j)) = \mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_j)$,
(4) $G_1(\mathsf{RH}(X_i)) \boxplus \mathsf{LH}(X_i) = G_1(\mathsf{RH}(X_k)) \boxplus \mathsf{LH}(X_k)$.

By accounting for three possibilities of the good node in $\mathcal{T}$, and six corresponding possibilities of the collision at round 1, in Case 4.2, Bad happens with probability at most $|\mathcal{D}| \cdot \frac{18}{M^6N^7}$.

Case 5: $|S| \in \{3, 4\}$ and $|S^*| = 6$. Thus at round one we have at least two collisions. Partition the six queries into three groups $\{X_s, X_i\}, \{X_j, X_k\}, \{X_\ell, X_r\}$. We consider the following cases.

Case 5.1: The two collisions at round 1 are between the same two groups, say the first and the second groups. From the characterization of collisions, the collisions at round 1 must be between X_i and X_k, and between X_s and X_j. This leads to the following equations of G_1 and G_4:

$$G_1(\mathsf{RH}(X_i)) \boxplus \mathsf{LH}(X_i) = G_1(\mathsf{RH}(X_k)) \boxplus \mathsf{LH}(X_k)$$
$$G_1(\mathsf{RH}(X_s)) \boxplus \mathsf{LH}(X_s) = G_1(\mathsf{RH}(X_j)) \boxplus \mathsf{LH}(X_j)$$
$$G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_k)) = \Big(\mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_k)\Big) \boxminus \Big(\mathsf{RH}(X_i) \boxminus \mathsf{RH}(X_k)\Big)$$
$$G_4(\mathsf{LH}(C_s)) \boxminus G_4(\mathsf{LH}(C_j)) = \Big(\mathsf{RH}(C_s) \boxminus \mathsf{RH}(C_j)\Big) \boxminus \Big(\mathsf{RH}(X_s) \boxminus \mathsf{RH}(X_j)\Big)$$

Since $\mathsf{LH}(C_k) = \mathsf{LH}(C_j)$ and $\mathsf{LH}(C_i) = \mathsf{LH}(C_s)$, the equations on G_4 imply that

$$\Big(\mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_k)\Big) \boxminus \Big(\mathsf{RH}(X_i) \boxminus \mathsf{RH}(X_k)\Big)$$
$$= \Big(\mathsf{RH}(C_s) \boxminus \mathsf{RH}(C_j)\Big) \boxminus \Big(\mathsf{RH}(X_s) \boxminus \mathsf{RH}(X_j)\Big) \ . \tag{36}$$

Let $\mathcal{D}_2$ be the subset of $\mathcal{D}$ such that any $(C_i, \ldots, C_s) \in \mathcal{D}_2$ satisfies Equation (36) above. Again

$$\frac{|\mathcal{D}_2|}{|\mathcal{D}|} \leq \frac{1}{N-2} \ .$$

Fix an arbitrary element of $\mathcal{D}_2$ for $(C_i, \ldots, C_s)$. In this case $R_1, \ldots, R_s$ are distinct. Since $|S| \geq 3$, there exists a subset $\mathcal{I} \subseteq \{i, j, k, \ell, r, s\}$ such that $|\mathcal{I}| = 3$ and the values L_t, with $t \in \mathcal{I}$, are distinct. Hence the chance that the following events happen is $\frac{1}{M^8N^4}$:

(1) $G_2(L_t) \boxplus \mathsf{RH}(X_t) = R_t$ for every $t \in \mathcal{I}$,
(2) $G_3(R_m) \boxplus L_m = \mathsf{LH}(C_m)$ for every $m \in \{i, j, k, \ell, r, s\}$,
(3) $G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_k)) = \Big(\mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_k)\Big) \boxminus \Big(\mathsf{RH}(X_i) \boxminus \mathsf{RH}(X_k)\Big)$,
(4) $G_1(\mathsf{RH}(X_i)) \boxplus \mathsf{LH}(X_i) = G_1(\mathsf{RH}(X_k)) \boxplus \mathsf{LH}(X_k)$,
(5) $G_1(\mathsf{RH}(X_s)) \boxplus \mathsf{LH}(X_s) = G_1(\mathsf{RH}(X_j)) \boxplus \mathsf{LH}(X_j)$.

By accounting for three possibilities of the two collisions at round 1, in Case 5.1, Bad happens with probability at most $|\mathcal{D}_2| \cdot \frac{3}{M^8N^4} \leq |\mathcal{D}| \cdot \frac{3}{(N-2)M^6N^6}$.

Case 5.2: The two collisions at round 1 are between different two groups. Assume that those collisions are between X_i and X_k, and between X_j and X_r; the other cases are similar. Fix an arbitrary element of $\mathcal{D}$ for $(C_i, \ldots, C_s)$. In this case $R_1, \ldots, R_s$ are distinct. Since $|S| \geq 3$, there exists a subset $\mathcal{I} \subseteq \{i, j, k, \ell, r, s\}$ such that $|\mathcal{I}| = 3$ and the values L_t, with $t \in \mathcal{I}$, are distinct. Hence the chance that the following events happen is $\frac{1}{M^8N^5}$:

(1) $G_2(L_t) \boxplus \mathsf{RH}(X_t) = R_t$ for every $t \in \mathcal{I}$,
(2) $G_3(R_m) \boxplus L_m = \mathsf{LH}(C_m)$ for every $m \in \{i, j, k, \ell, r, s\}$,
(3) $G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_k)) = \Big(\mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_k)\Big) \boxminus \Big(\mathsf{RH}(X_i) \boxminus \mathsf{RH}(X_k)\Big)$,
(4) $G_4(\mathsf{LH}(C_j)) \boxminus G_4(\mathsf{LH}(C_r)) = \Big(\mathsf{RH}(C_j) \boxminus \mathsf{RH}(C_r)\Big) \boxminus \Big(\mathsf{RH}(X_j) \boxminus \mathsf{RH}(X_r)\Big)$,
(5) $G_1(\mathsf{RH}(X_i)) \boxplus \mathsf{LH}(X_i) = G_1(\mathsf{RH}(X_k)) \boxplus \mathsf{LH}(X_k)$,
(6) $G_1(\mathsf{RH}(X_r)) \boxplus \mathsf{LH}(X_r) = G_1(\mathsf{RH}(X_j)) \boxplus \mathsf{LH}(X_j)$.

By accounting for three possibilities of the two collisions at round 1, in Case 5.2, Bad happens with probability at most $|\mathcal{D}| \cdot \frac{3}{M^8N^5} \leq |\mathcal{D}| \cdot \frac{3}{M^7N^6}$.

Case 6: $|S| = 4$ and $|S^*| = 5$. This is similar to Case 5 and the chance that Bad happens in this case is at most

$$3\Big(|\mathcal{D}| \cdot \frac{3}{(N-2)M^6N^6} + |\mathcal{D}| \cdot \frac{3}{M^7N^6}\Big) = |\mathcal{D}| \cdot \frac{9}{(N-2)M^6N^6} + |\mathcal{D}| \cdot \frac{9}{M^7N^6} \ ,$$

where the factor 3 is due to the accounting of the three possibilities of the good node of $\mathcal{T}$.

Case 7: $|S| = 3$ and $|S^*| = 5$. Since $|S^*| = 5$, there must be a good node, say $v_{i,j}$ in $\mathcal{T}$. On the other hand, since $|S| = 3$, there must be exactly three collisions at round 1. Assume that those collisions are between X_i and X_ℓ, between X_j and X_s, and between X_k and X_r; the other cases are similar. The collisions between X_i and X_ℓ, and between X_j and X_s lead to the following equations:

$$G_1(\mathsf{RH}(X_i)) \boxplus \mathsf{LH}(X_i) = G_1(\mathsf{RH}(X_\ell)) \boxplus \mathsf{LH}(X_\ell)$$
$$G_1(\mathsf{RH}(X_s)) \boxplus \mathsf{LH}(X_s) = G_1(\mathsf{RH}(X_j)) \boxplus \mathsf{LH}(X_j)$$
$$G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_\ell)) = \Big(\mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_\ell)\Big) \boxminus \Big(\mathsf{RH}(X_i) \boxminus \mathsf{RH}(X_\ell)\Big)$$
$$G_4(\mathsf{LH}(C_s)) \boxminus G_4(\mathsf{LH}(C_j)) = \Big(\mathsf{RH}(C_s) \boxminus \mathsf{RH}(C_j)\Big) \boxminus \Big(\mathsf{RH}(X_s) \boxminus \mathsf{RH}(X_j)\Big)$$

On the other hand, the collision between X_i and X_j at round 2 leads to another equation of G_4:

$$G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_j)) = \mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_j) \ .$$

Since $\mathsf{LH}(C_s) = \mathsf{LH}(C_i)$, the three equations above on G_4 imply that

$$\Big(\mathsf{RH}(C_s) \boxminus \mathsf{RH}(C_j)\Big) \boxminus \Big(\mathsf{RH}(X_s) \boxminus \mathsf{RH}(X_j)\Big) = \mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_j) \ . \tag{37}$$

Let $\mathcal{D}_3$ be the subset of $\mathcal{D}$ such that any $(C_i, \ldots, C_s) \in \mathcal{D}_3$ satisfies Equation (37) above. Again,

$$\frac{|\mathcal{D}_3|}{|\mathcal{D}|} \leq \frac{1}{N-2} \ .$$

Fix an arbitrary element of $\mathcal{D}_3$ for $(C_i, \ldots, C_s)$. In this case, $R_j, \ldots, R_s$ are distinct, and L_i, L_j, L_k are distinct. Hence the chance that the following events happen is $\frac{1}{M^7N^5}$:

(1) $G_2(L_t) \boxplus \mathsf{RH}(X_t) = R_t$ for every $t \in \{i, j, k\}$,
(2) $G_3(R_m) \boxplus L_m = \mathsf{LH}(C_m)$ for every $m \in \{j, k, \ell, r, s\}$,
(3) $G_1(\mathsf{RH}(X_i)) \boxplus \mathsf{LH}(X_i) = G_1(\mathsf{RH}(X_\ell)) \boxplus \mathsf{LH}(X_\ell)$,
(4) $G_1(\mathsf{RH}(X_s)) \boxplus \mathsf{LH}(X_s) = G_1(\mathsf{RH}(X_j)) \boxplus \mathsf{LH}(X_j)$,
(5) $G_4(\mathsf{LH}(C_i)) \boxminus G_4(\mathsf{LH}(C_\ell)) = \Big(\mathsf{RH}(C_i) \boxminus \mathsf{RH}(C_\ell)\Big) \boxminus \Big(\mathsf{RH}(X_i) \boxminus \mathsf{RH}(X_\ell)\Big)$,
(6) $G_4(\mathsf{LH}(C_s)) \boxminus G_4(\mathsf{LH}(C_j)) = \Big(\mathsf{RH}(C_s) \boxminus \mathsf{RH}(C_j)\Big) \boxminus \Big(\mathsf{RH}(X_s) \boxminus \mathsf{RH}(X_j)\Big)$.

By accounting for four possibilities of the three collisions at round 1, and three possibilities of the good node in $\mathcal{T}$, in this case, Bad happens with probability at most $|\mathcal{D}_3| \cdot \frac{12}{(N-2)M^7N^5} \leq |\mathcal{D}| \cdot \frac{12}{(N-2)M^6N^6}$.

Summing over all cases, we obtain

$$\Pr[\mathsf{Bad}] \leq |\mathcal{D}| \cdot \left(\frac{4}{M^5N^7} + \frac{33}{(N-2)M^6N^6} + \frac{39}{M^6N^7}\right) \ . \tag{38}$$

Wrapping up. From Equation (33) and Equation (38), we conclude that

$$\Pr[\mathsf{Bad}] \leq \Pr[\mathsf{Good}] \cdot \frac{N}{M-9}\Big(\frac{4}{M} + \frac{33N}{M^2(N-2)} + \frac{39}{M^2}\Big) \ .$$

This concludes the proof.