A Note on the Chi-square Method : A Tool for Proving Cryptographic Security

Srimanta Bhattacharya, Mridul Nandi

Indian Statistical Institute, Kolkata, India.

Abstract. Very recently (in CRYPTO 2017) Dai, Hoang, and Tessaro have introduced the *Chi-square method* (χ^2 method) which can be applied to obtain an upper bound on the statistical distance between two joint probability distributions. The authors have applied this method to prove the *pseudorandom function security* (PRF-security) of sum of two random permutations. In this work, we revisit their proof and find a non-trivial gap in the proof. We plug this gap for two specific cases and state the general case as an assumption whose proof is essential for the completeness of the proof by Dai *et al.*. A complete, correct, and transparent proof of the full security of the sum of two random permutations construction is much desirable, especially due to its importance and two decades old legacy. The proposed χ^2 method seems to have potential for application to similar problems, where a similar gap may creep into a proof. These considerations motivate us to communicate our observation in a formal way.
On the positive side, we provide a very simple proof of the PRF-security of the *truncated random permutation* construction (a method to construct PRF from a random permutation) using the χ^2 method. We note that a proof of the PRF-security due to Stam is already known for this construction in a purely statistical context. However, the use of the χ^2 method makes the proof much simpler.

Keywords. random permutation, pseudorandom function, χ^2-distance, KL divergence, total variation distance, Pinsker's inequality, sum of random permutation, truncated random permutation.

1 Introduction

Different tools from probability and statistics are now heavily used in different areas in cryptography. In this paper, we focus on a statistical tool, termed χ^2 method, which has been introduced by Dai, Hoang, and Tessaro in CRYPTO 2017 ([DHT17a]). Although a method which is essentially similar to the χ^2 method is known in statistics (since 1978), we believe that the χ^2 method is new in the context of cryptography. In [DHT17a], this method has been used to show pseudorandom function security (PRF-security) of two well known constructions, namely sum of random permutations ([Pat08b,Pat10,BI99,Luc00]) and encrypted Davis-Meyer (EDM) ([CS16,MN17]). Further, we feel that this method may help us to obtain tight (and simplified) proofs for certain constructions where proofs so far have evaded more classical methods, such as the H-coefficient method ([Pat08a]).

χ^2 METHOD. The *distinguishing advantage* of a family of keyed functions is bounded by the total variation (also known as *statistical distance*) between the output distribution of the family and the output distribution of a random function. Total variation between two probability distributions $\mathbf{P_0}$ and $\mathbf{P_1}$ over a sample space Ω, denoted $d_{\mathrm{TV}}(\mathbf{P_0}, \mathbf{P_1})$, is defined as the half of L_1-norm $\|\mathbf{P_0} - \mathbf{P_1}\|_1 := \sum_{x \in \Omega} |\mathbf{P_0}(x) - \mathbf{P_1}(x)|$. In [DHT17a], the authors have revisited a variation of the additivity property of the KL divergence between two joint distributions. The authors have termed it χ^2 method. When $\mathbf{P_0}$ and $\mathbf{P_1}$ are joint distributions, this method provides an upper bound on $\|\mathbf{P_0} - \mathbf{P_1}\|_1$ based on the χ^2-distances between the conditional distributions of $\mathbf{P_0}$ and $\mathbf{P_1}$. Next, we recall the definition of χ^2-distance. In what follows, we use the convention that $0/0 = 0$.

Definition 1. *The χ^2-distance between distributions $\mathbf{P_0}$ and $\mathbf{P_1}$ (over a sample space Ω) with $\mathbf{P_0} \ll \mathbf{P_1}$ (i.e., the support of $\mathbf{P_0}$ is contained in the support of $\mathbf{P_1}$) is defined as*

$$d_{\chi^2}(\mathbf{P_0}, \mathbf{P_1}) := \sum_{x \in \Omega} \frac{(\mathbf{P_0}(x) - \mathbf{P_1}(x))^2}{\mathbf{P_1}(x)}.$$

χ^2-distance has its origin in mathematical statistics dating back to Pearson (see [LV87] for some history). It can be seen that χ^2-distance is not symmetric and hence it is not a metric. However, this is useful for bounding other metrics, e.g., total variation. In the following, we briefly describe the χ^2 method (see Section A for details and proof).

Let $\mathsf{X} = (\mathsf{X}_1, \ldots, \mathsf{X}_q)$ and $\mathsf{Y} = (\mathsf{Y}_1, \ldots, \mathsf{Y}_q)$ be two multivariate random variables taking values from Ω^q. In order to simplify the notation, we denote by X^{i-1} the joint random variable $(\mathsf{X}_1, \ldots, \mathsf{X}_{i-1})$. Let $\mathbf{P_0}_{x_1,\ldots,x_{i-1}}$ denote the conditional probability distribution of X_i given $\mathsf{X}_1 = x_1, \ldots, \mathsf{X}_{i-1} = x_{i-1}$. We similarly write $\mathbf{P_1}_{x_1,\ldots,x_{i-1}}$ for the distribution of Y_i given $\mathsf{Y}_1 = x_1, \ldots, \mathsf{Y}_{i-1} = x_{i-1}$. Then the χ^2 method says

$$d_{\mathrm{TV}}(\mathsf{X}, \mathsf{Y}) \leq \left(\frac{1}{2} \sum_{i=1}^{q} \mathbf{Ex}[\chi^2(\mathsf{X}_1, \ldots, \mathsf{X}_{i-1})] \right)^{\frac{1}{2}}, \tag{1}$$

where $\chi^2(x_1, \ldots, x_{i-1}) = d_{\chi^2}(\mathbf{P_0}_{x_1,\ldots,x_{i-1}}, \mathbf{P_1}_{x_1,\ldots,x_{i-1}})$ and for all $x_1, \ldots, x_{i-1}$, $\mathbf{P_0}_{x_1,\ldots,x_{i-1}} \ll \mathbf{P_1}_{x_1,\ldots,x_{i-1}}$. Note that we need this condition to define d_{χ^2}.

XOR OF TWO RANDOM PERMUTATIONS. *XOR or sum of two random permutations* is a well known construction, proposed and studied by Hall *et al.* in [HWKS98], for conversion of *pseudorandom permutations* (PRPs) into *pseudorandom functions* (PRFs) [1]. Given a permutation $\pi : \{0,1\}^n \mapsto \{0,1\}^n$, the

[1] This line of work was initiated by Bellare *et al.* in [BKR98] who coined the term "Luby-Rackoff backwards" for such conversion.

construction creates a function $f : \{0,1\}^{n-1} \to \{0,1\}^n$, defined as $f(x) = \pi(0||x) \oplus \pi(1||x)$. When π is chosen uniformly at random from Perm_n, the set of all permutations of $\{0,1\}^n$, how well does f resemble (in a certain well defined sense) a random function with the same domain and range (a function chosen uniformly from the set of all functions from the domain to the range)? A satisfactory answer to this question remained elusive for over two decades. There have been attempts ([Luc00,BI99,Pat08b,Pat10]) to prove *information-theoretic security* of the construction. However, the proofs either fell short of proving *full security* (to be made precise in the next section) of the construction([Luc00]) or were sketchy ([BI99]) or contained non-trivial gaps and were difficult to follow ([Pat08b,Pat10]) as has also been observed by the authors of [DHT17a].[2] Also, as a related problem, Cogliati, Lampe, and Patarin [CLP14] gave weaker bounds for the case of the sum of at least three permutations. The XOR construction is important since it has been used to obtain some constructions achieving beyond birthday (or sometimes almost full) security (e.g., CENC [Iwa06], PMAC_Plus [BR02] and ZMAC [IMPS17]).

1.1 Main Results in the Paper

In [DHT17a], Dai *et al.* have used the χ^2 method to prove full security of the XOR construction (XOR of two random permutations). In this paper, we have a closer inspection of the proof and we find a non-trivial gap in it. The gap is due to incorrect equalities involving conditional expectations. We have also made an attempt to fix the proof and we have shown that the proof can be fixed under Asssumption 1 (stated in Section 4). We have proved the assumption for two special cases and left the general case as an open problem.

In this note, we communicate the above observation formally. This serves two purposes:(a) to motivate a flawless proof of this problem, especially owing to its importance and a two-decades old legacy, (b) to prevent these types of loopholes from creeping into the proofs involving the χ^2 method, especially since the method seems to have potential for application to similar problems.

Truncation of Random Permutation. Although the application (in [DHT17a]) of the χ^2 method to the XOR construction contains gap, this technique can be powerful for bounding PRF-security of other constructions. In fact, in [DHT17a], the authors applied this method to bound the PRF-security of the EDM (or encrypted Davis-Meyer) construction. In this note, we apply this technique to the **truncated random permutation** construction and obtain a very simple proof of the known tight bound on the PRF-security of the construction. This has been studied by Stam (in a statistical context) in 1978 [Sta78] and later by many others (e.g., [GG15,GG16,GGM17,HWKS98,BI99]). Stam's proof technique is very close to the χ^2 method. However, the other proofs are very different and produce different results. The difference between the proof methods of the relevant results from [HWKS98], [BI99], [GG15] and [Sta78] is discussed in [GGM17]. Our

[2] A quote from the paper [DHT17a] "`Patarin's tight proof is very involved, with some claims remaining open or unproved.`"

proof approach is more modular and uses the χ^2 method explicitly. We discuss these very briefly in Remark 1 and Remark 2

The PRF property of the truncated random permutation construction has recently been used in the key derivation for the AES-GCM, Counter based authenticated encryption constructions [GLL17].

1.2 Organization of the Paper

The rest of the paper is organized as follows. In the next section, we provide a brief overview of relevant security notions and the χ^2 method. There we also discuss the two constructions: XOR of two random permuations construction and trucated random permutation construction. Section 3 is devoted to the proof of Theorem 2. In Section 4, we discuss the proof, by Dai *et al.*, of the full security of the XOR of two random permutation construction, where we also point out the gap in it. In Section 5, we provide the proofs of Assumption A1 for two specific cases. We conclude in Section 6 by remarking on an identical gap in the proof of full security of a related construction, termed *XOR of two independent random permutations*. Finally, in Appendix A, we provide a self-contained proof of the χ^2 method; essential ingredients of the proof is same as that of [DHT17a], however, we also cover the finer details (such as the proof of the Pinsker's inequality).

2 Preliminaries

Notation and Convention. We use the short-hand notation X^t to denote a tuple $(X_1, \ldots, X_t)$. We also write $\mathcal{S}^t$ to denote the t-fold Cartesian product of the set $\mathcal{S}$ with itself. It will be clear from the context whether X^t means a t-tuple (when X is a tuple) or product set (when X is a set).

We use notations $\mathsf{X}, \mathsf{Y}, \mathsf{Z}$ etc. (possibly with suffix) to represent random variables over some sets. Following the above notational convention, X^t would represent a t-tuple of random variables or random vector $(\mathsf{X}_1, \ldots, \mathsf{X}_t)$. We use $\mathcal{E}, \mathcal{S}, \mathcal{T}$ etc. (possibly with suffix) to denote sets. $\mathcal{A}$ will always represent an adversary.

In this paper, we fix a positive integer n, and we denote 2^n by N.

2.1 PRF-Security Definition

Pseudorandom function (PRF) is a very popular security notion in cryptography. While analyzing a message authentication code (MAC), we mostly study its PRF-security as it is a stronger notion than MAC. It has also been used to define encryption schemes, authenticated encryptions, and other cryptographic algorithms.

Now we formally define the *PRF-advantage* of an algorithm or a keyed function. By $\mathsf{X} \leftarrow_{\$} \mathcal{S}$ we mean that X is sampled uniformly from a finite set $\mathcal{S}$. Let m and p be positive integers. Let RP_m denote the random permutation chosen uniformly from Perm_m, the set of all permutations on $\{0,1\}^m$, i.e.,

$\mathsf{RP}_m \leftarrow_\$ \mathsf{Perm}_m$. Similarly, let $\mathsf{RF}_{m\to p} \leftarrow_\$ \mathsf{Func}_{m\to p}$ (the set of all functions from $\{0,1\}^m$ to $\{0,1\}^p$). Let $\mathcal{K}$ be a finite set (it is the key space of the construction). Given a function $f : \mathcal{K} \times \{0,1\}^m \to \{0,1\}^p$ and for every $k \in \mathcal{K}$, we denote f_k to represent the function (also called keyed function) $f(k,\cdot) \in \mathsf{Func}_{m\to p}$. We now define the PRF-advantage of an oracle adversary $\mathcal{A}$ against f as follows.

Definition 2 (PRF-advantage). *Let $\mathcal{A}$ be a distinguisher (oracle algorithm) and $f : \mathcal{K} \times \{0,1\}^m \to \{0,1\}^p$. Then, the PRF-advantage of $\mathcal{A}$ against f is defined as*

$$\mathbf{Adv}_f^{\mathrm{prf}}(\mathcal{A}) = |\mathbf{P}[\mathcal{A}^{f_K} \to 1 \ : \ K \leftarrow_\$ \mathcal{K}] - \mathbf{P}[\mathcal{A}^{RF_{m\to p}} \to 1]|.$$

As we restrict to only deterministic keyed functions (i.e., functions which give same output on same input) we can assume, without loss of generality, that the adversary does not repeat its queries. In other words, if $\mathsf{Q}_1,\ldots,\mathsf{Q}_q$ are all queries then these are distinct. We can also assume that $\mathcal{A}$ is deterministic as it can always run with the best random coins which maximize the advantage. Suppose $\mathcal{A}$ makes q distinct queries adaptively, denoted $\mathsf{Q}_1,\ldots,\mathsf{Q}_q$, and obtains responses $\mathsf{U}_1,\ldots,\mathsf{U}_q$. So, when $\mathcal{A}$ in interacting with $\mathsf{RF}_{m\to p}$, the outputs are uniformly and independently distributed over $\{0,1\}^p$ which we denote as $\mathsf{U}_1,\ldots,\mathsf{U}_q \leftarrow_\$ \{0,1\}^p$.

Similarly, let $\mathsf{X}_1,\ldots,\mathsf{X}_q$ denote the outputs of f_{K} where $\mathsf{K} \leftarrow_\$ \mathcal{K}$. We denote the probability distributions associated with $\mathsf{U}_1,\ldots,\mathsf{U}_q$ and $\mathsf{X}_1,\ldots,\mathsf{X}_q$ by $\mathbf{P_1}$ and $\mathbf{P_0}$ respectively. Thus,

$$\mathbf{Adv}_f^{\mathrm{prf}}(\mathcal{A}) = |\mathbf{P_1}(\mathcal{E}) - \mathbf{P_0}(\mathcal{E})| \tag{2}$$

where $\mathcal{E}$ is the set of all q-tuple of responses $x^q := (x_1,\ldots,x_q) \in (\{0,1\}^n)^q$ for which $\mathcal{A}$ returns 1. From the definition the total variation (also known as the *statistical distance*) between $\mathbf{P_0}$ and $\mathbf{P_1}$ is

$$d_{\mathrm{TV}}(\mathbf{P_0},\mathbf{P_1}) \stackrel{\mathrm{def}}{=} \frac{1}{2}\sum_{x^q\in(\{0,1\}^n)^q} |\mathbf{P_0}(x^q) - \mathbf{P_1}(x^q)| = \max_{\mathcal{E}\subseteq\Omega}(\mathbf{P_0}(\mathcal{E}) - \mathbf{P_1}(\mathcal{E})). \tag{3}$$

Hence,

$$\mathbf{Adv}_f^{\mathrm{prf}}(\mathcal{A}) \le d_{\mathrm{TV}}(\mathbf{P_1},\mathbf{P_0}).^3$$

Thus, the main cryptographic objective (that of determining the PRF-advantage $\mathbf{Adv}_f^{\mathrm{prf}}(\mathcal{A})$) turns out to be a purely probability or statistical problem. Next, we discuss the χ^2 method which provides an upper bound of total variation between two joint distributions.

2.2 χ^2 Method

Let $\mathsf{X} := (\mathsf{X}_1,\ldots,\mathsf{X}_q)$ and $\mathsf{Z} := (\mathsf{Z}_1,\ldots,\mathsf{Z}_q)$ are two random vectors of size q distributed over Ω^q. Let us denote the probability distributions of X and Z as

[3] In fact, in this setting, i.e, for information theoretic security, there always exists an adversary $\mathcal{A}'$ such that $\mathbf{Adv}_f^{\mathrm{prf}}(\mathcal{A}') = d_{\mathrm{TV}}(\mathbf{P_1},\mathbf{P_0})$; $\mathcal{A}'$ returns 1 for any $x^q \in \mathcal{E}'$, where $\mathcal{E}'$ is such that $d_{\mathrm{TV}}(\mathbf{P_1},\mathbf{P_0}) = \mathbf{P_0}(\mathcal{E}') - \mathbf{P_1}(\mathcal{E}')$.

$\mathbf{P_0}$ and $\mathbf{P_1}$ respectively. We denote the conditional probability distributions as follows.

$$\mathbf{P}_{\mathbf{0}|x^{i-1}}(x_i) = \mathbf{P}(\mathsf{X}_i = x_i|\mathsf{X}_1 = x_1, \ldots, \mathsf{X}_{i-1} = x_{i-1})$$
$$\mathbf{P}_{\mathbf{1}|x^{i-1}}(x_i) = \mathbf{P}(\mathsf{Z}_i = x_i|\mathsf{Z}_1 = x_1, \ldots, \mathsf{Z}_{i-1} = x_{i-1})$$

When $i = 1$, $\mathbf{P}_{\mathbf{0}|x^{i-1}}(x_1)$ represents $\mathbf{P}(\mathsf{X}_1 = x_1)$. Similarly, for $\mathbf{P}_{\mathbf{1}|x^{i-1}}(x_1)$. Let $x^{i-1} \in \Omega^{i-1}$, $i \geq 1$. Let us denote the χ^2-distance between $\mathbf{P}_{\mathbf{0}|x^{i-1}}$ and $\mathbf{P}_{\mathbf{1}|x^{i-1}}$ as $\chi^2(x^{i-1})$, i.e.,

$$\chi^2(x^{i-1}) := d_{\chi^2}(\mathbf{P}_{\mathbf{0}|x^{i-1}}, \mathbf{P}_{\mathbf{1}|x^{i-1}}).$$

Thus, χ^2 is a real valued function. The next theorem is the crux of the χ^2 method; it bounds the total variation between two joint distributions in terms of the χ^2-distance between the corresponding conditional distributions.

Theorem 1 ([DHT17a]). *Suppose* $\mathbf{P_0}$ *and* $\mathbf{P_1}$ *denote probability distributions of* $\mathsf{X} := (\mathsf{X}_1, \ldots, \mathsf{X}_q)$ *and* $\mathsf{Z} := (\mathsf{Z}_1, \ldots, \mathsf{Z}_q)$ *and for all* $x_1, \ldots, x_{i-1}$*, we have* $\mathbf{P}_{\mathbf{0}|x^{i-1}} \ll \mathbf{P}_{\mathbf{1}|x^{i-1}}$*. Then*

$$d_{\mathrm{TV}}(\mathbf{P_0}, \mathbf{P_1}) \leq \left(\frac{1}{2} \sum_{i=1}^{q} \mathbf{Ex}[\chi^2(\mathsf{X}^{i-1})] \right)^{\frac{1}{2}}.$$

For the sake of completeness, we provide a complete proof of this theorem in Appendix A. In our setup, note that $\mathsf{Z}_1, \ldots, \mathsf{Z}_q \leftarrow_{\$} \{0,1\}^p$ for some p and hence $\mathbf{P}_{\mathbf{1}|x^{i-1}}(x_i) = \frac{1}{2^p}$ for all x^i. So,

$$\mathbf{Ex}[\chi^2(\mathsf{X}^{i-1})] = 2^p \sum_{x_i} \mathbf{Ex}_{\mathsf{X}^{i-1}} \left[\left(\mathbf{P}(\mathsf{X}_i = x_i|\mathsf{X}_1, \ldots, \mathsf{X}_{i-1}) - \frac{1}{2^p} \right)^2 \right].$$

In the following subsection, we describe two constructions for which this method was applied.

2.3 Two Random Permutation Based Constructions

In this paper, we mainly deal with two constructions based on a random permutation RP_n. Similar to a random function, if all queries to a random permutation RP_n are distinct and depends only on the previous responses (which is the case for an adversary), the outputs $\mathsf{V}_1, \ldots, \mathsf{V}_q$ behave like a random sample without replacement (WOR) from $\{0,1\}^n$. We write $\mathsf{V}_1, \ldots, \mathsf{V}_q \leftarrow_{\mathrm{wor}} \{0,1\}^n$ to denote this. More formally, for all *distinct* $x_1, \ldots, x_q \in \{0,1\}^n$, $\mathbf{P}(\mathsf{V}_1 = x_1, \ldots \mathsf{V}_q = x_q) = \frac{1}{(N)_q}$, where $(N)_q = N(N-1)\cdots(N-q+1)$. Now, we briefly describe the constructions.

(1) XOR **Construction.** Define $\mathsf{XOR}_\pi : \{0,1\}^{n-1} \to \{0,1\}^n$ to be the construction that takes a permutation $\pi \in \mathsf{Perm}_n$ as a key, and on input $x \in \{0,1\}^{n-1}$

it returns $\pi(x\|0) \oplus \pi(x\|1)$. Thus, XOR construction based on a random permutation RP_n returns $\mathsf{X}_1, \ldots, \mathsf{X}_q$ where $\mathsf{X}_1 := \mathsf{V}_1 \oplus \mathsf{V}_2$, ..., $\mathsf{X}_q := \mathsf{V}_{2q-1} \oplus \mathsf{V}_{2q}$ and $\mathsf{V}_1, \ldots, \mathsf{V}_{2q} \leftarrow_{\mathrm{wor}} \{0,1\}^n$.

(2) trRP **Construction.** Let $m \leq n$ and trunc_m denotes the *truncation function* which returns the first m bits of $x \in \{0,1\}^n$. Truncated random permutation is a composition of random permutation followed by a truncation function. More formally, we define for every $x \in \{0,1\}^n$,

$$\mathsf{trRP}_m(x) = \mathsf{trunc}_m(\mathsf{RP}_n(x)).$$

Note that it is a function family, keyed by random permutation, mapping the set of all n-bit sequences to the set of all m-bit sequences. Let $\mathsf{X}_1, \ldots, \mathsf{X}_q$ denote the q outputs of trRP_m. Then $\mathsf{X}_i = \mathsf{trunc}_m(\mathsf{V}_i)$ for all i.

PRF-security of this construction has been studied by Stam in 1978, though in a much broader context (see [Sta78] for details), and later by others (e.g., [HWKS98,BI99,GG15,GG16,GGM17]). In particular, Stam proved the following statement.

Theorem 2 ([Sta78]). *Let* $\mathsf{V}_1, \ldots, \mathsf{V}_q \leftarrow_{\mathrm{wor}} \{0,1\}^n$, $\mathsf{U}_1, \ldots, \mathsf{U}_q \leftarrow_{\$} \{0,1\}^m$ *and* $\mathsf{X}_i = \mathsf{trunc}_m(\mathsf{V}_i)$ *for all* i. *Then*

$$d_{\mathrm{TV}}(\mathsf{X}, \mathsf{U}) \leq \frac{1}{2}\left(\frac{(M-1)q(q-1)}{(N-1)(N-q+1)}\right)^{\frac{1}{2}}$$

where $\mathsf{X} = (\mathsf{X}_1, \ldots, \mathsf{X}_q)$ *and* $\mathsf{U} = (\mathsf{U}_1, \ldots, \mathsf{U}_q)$.

The following corollary (though not proved by Stam) is immediate from the relationship between PRF-advantage and total variation.

Corollary 1. *Let* $M = 2^m$, $N = 2^n$ *and* $m \leq n$. *For any adversary* $\mathcal{A}$ *making* q *queries we have*

$$\mathbf{Adv}^{\mathrm{prf}}_{\mathsf{trRP}_m}(\mathcal{A}) \leq \frac{1}{2}\left(\frac{(M-1)q(q-1)}{(N-1)(N-q+1)}\right)^{\frac{1}{2}}.$$

Remark 1. The upper bounds on the PRF-advantage of the trRP Construction given in [HWKS98,GG15] are different (and weaker) than the one obtained by Stam. Although the bounds are similar for some choices of parameters. In [GGM17], all these results are mentioned, and the proofs are briefly surveyed. In [GGM17], a general tight lower bound on the PRF-advantage has been proved (improving on the lower bound declared in [HWKS98]).

3 Proof of Theorem 2 Using the χ^2 Method

Now we provide an alternative proof of Theorem 2 using the χ^2 method. We briefly recall the setup. Here $\mathsf{V}_1, \ldots, \mathsf{V}_q \leftarrow_{\mathrm{wor}} \{0,1\}^n$ and $\mathsf{X}_i = \mathsf{trunc}_m(\mathsf{V}_i)$. Let

$x \in \{0,1\}^m$, $i \geq 1$ be an integer, and $K = N/M$. Also, let H denote the number of $j < i$, for which $\mathsf{trunc}_m(\mathsf{V}_j) = x$. The probability distribution of H is well known as the hypergeometric distribution $\mathsf{HG}(N, K, i-1)$. For every $\max(0, s+K-N) \leq a \leq \min(K, s)$ we have

$$\mathbf{P}(\mathsf{H} = a) = \frac{\binom{K}{a} \times \binom{N-K}{s-a}}{\binom{N}{s}}.$$

The following fact states the expectation and variance formula of a hypergeometric distribution. Its proof can be found in standard probability theory text books and hence we skip it.

Fact 1 *Let* H *follow hypergeometric distribution* $\mathsf{HG}(N, K, s)$ *and let* p *denote* $\frac{K}{N}$. *Then,*

$$\mathbf{Ex}[\mathsf{H}] = sp. \tag{4}$$

$$\mathbf{Var}[\mathsf{H}] := \mathbf{Ex}[\mathsf{H} - \mathbf{Ex}[\mathsf{H}]]^2 = sp(1-p) \times \frac{N-s}{N-1}. \tag{5}$$

As an aside, we mention that the factor $\frac{N-s}{N-1}$ is also known as the finite sampling correction factor. Up to this factor, the expression of variance is same as that of the binomial distribution.

Now, we apply the χ^2 method to bound the total variation $d_{\mathrm{TV}}(\mathsf{X}, \mathsf{U})$, where $\mathsf{U}_1, \ldots, \mathsf{U}_q \leftarrow_{\$} \{0,1\}^m$. Let $\mathbf{P_0}$ and $\mathbf{P_1}$ denote the probability distributions of X and U respectively. Note that

$$\begin{aligned}
\mathbf{P}_{\mathbf{0}|x^{i-1}}(x) &= \mathbf{P}(\mathsf{X}_i = x \mid \mathsf{X}_1 = x_1, \ldots, \mathsf{X}_{i-1} = x_{i-1}) \\
&= \mathbf{P}(\mathsf{V}_i \notin \mathcal{S}), \quad \text{where } \mathcal{S}_{i,x}(x^{i-1}) = \{v \in \{0,1\}^n : \exists j < i \ \mathsf{trunc}_m(v) = x_j\} \\
&= \frac{\frac{N}{M} - |\mathcal{S}_{i,x}(x^{i-1})|}{N - i + 1}.
\end{aligned}$$

Let $N_{i,x}(x^{i-1}) := |\mathcal{S}_{i,x}(x^{i-1})|$ and $\mathsf{H}_{i,x} = N_{i,x}(\mathsf{X}^{i-1})$. Then it is easy to see from the definition of the heypergeometric distribution that $\mathsf{H}_{i,x}$ follows $\mathsf{HG}(N, N/M, (i-1))$. Now, we compute the χ^2 function evaluated at x^{i-1}.

$$\begin{aligned}
\chi^2(x^{i-1}) &= \sum_{x \in [M]} M \left(\frac{\frac{N}{M} - N_{i,x}(x^{i-1})}{N - i + 1} - \frac{1}{M} \right)^2 \\
&= \sum_{x \in [M]} \frac{M}{(N-i+1)^2} \times \left(N_{i,x}(x^{i-1}) - \frac{i-1}{M} \right)^2.
\end{aligned}$$

Hence,

$$\mathbf{Ex}[\chi^2(\mathsf{X}^{i-1})] = \mathbf{Ex}\left[\sum_{x \in [M]} \frac{M}{(N-i+1)^2} \times \left(\mathsf{H}_{i,x} - \frac{i-1}{M} \right)^2 \right]$$

$$= \sum_{x \in [M]} \frac{M}{(N-i+1)^2} \times \mathbf{Var}[\mathsf{H}_{i,x}]. \tag{6}$$

This follows from the linearity of the expectation and the fact that $\mathbf{Ex}[\mathsf{H}_{i,x}] = (i-1)/M$. By substituting the value of $\mathbf{Var}[N_x]$ as described in the Fact 1, we obtain

$$\begin{aligned}\mathbf{Ex}[\chi^2(\mathsf{X}^{i-1})] &= \frac{M^2}{(N-i+1)^2} \times \frac{i-1}{M} \times \left(1 - \frac{1}{M}\right) \times \frac{N-i+1}{N-1} \\ &= \frac{(M-1)(i-1)}{(N-1)(N-i+1)}.\end{aligned}$$

Now by using Theorem 1 we have

$$\begin{aligned}d_{\mathrm{TV}}(\mathbf{P_0}, \mathbf{P_1}) &\leq \left(\frac{1}{2}\sum_{i=1}^{q} \mathbf{Ex}[\chi^2(\mathsf{X}^{i-1})]\right)^{\frac{1}{2}} \\ &= \left(\frac{1}{2}\sum_{i=1}^{q} \frac{(M-1)(i-1)}{(N-1)(N-i+1)}\right)^{\frac{1}{2}} \\ &\leq \left(\frac{1}{2}\sum_{i=1}^{q} \frac{(M-1)(i-1)}{(N-1)(N-q+1)}\right)^{\frac{1}{2}} \\ &= \frac{1}{2}\left(\frac{(M-1)q(q-1)}{(N-1)(N-q+1)}\right)^{\frac{1}{2}}. \qquad \square\end{aligned}$$

Remark 2. In order to draw comparison between our proof (using the χ^2 method) of Theorem 2 and the proof due to Stam, we remark that the main ideas of both the proofs are same; namely both use the chain rule of the KL divergence, concavity of the logarithm function, and also the hypergeometric distribution. However, unlike in our case (in (6)) Stam did not make explicit use of variance of the hypergeometric distribution. Instead, he used Jensen's inequality. Moreover, our proof is simpler and modular compared to Stam's proof with a more direct approach.

4 Overview of the Proof by Dai *et al.* and its Flaw

In this section, we provide a brief overview of the proof by Dai *et al.* to precisely point out the gap in their proof. In order to better emphasize, we provide a brief sketch of the proof due to Dai et al. We mostly follow the notation by the authors along with our notational convention. For example, we mostly use N instead of 2^n. Moreover, for simplicity we write the set $\{0,1\}^n \setminus \{0^n\}$ as $[N]^*$.

Theorem 3 ([DHT17a]). *Fix an integer $n \geq 8$ and let $N = 2^n$. For any adversary $\mathcal{A}$ that makes $q \leq \frac{N}{32}$ queries we have*

$$\mathbf{Adv}^{\mathrm{prf}}_{\mathsf{XOR}}(\mathcal{A}) \leq \frac{1.5q + 3\sqrt{q}}{N}.$$

Proof due to Dai et al in [DHT17a]. Let $\mathcal{A}$ be an adversary making exactly q distinct queries adaptively. As we have observed before, the output distributions of random function and XOR function do not depend on $\mathcal{A}$. In fact, $\mathsf{U}'_1, \ldots, \mathsf{U}'_q \leftarrow_{\$} \{0,1\}^n$ and $\mathsf{X}_1 := \mathsf{V}_1 \oplus \mathsf{V}_2, \ldots, \mathsf{X}_q := \mathsf{V}_{2q-1} \oplus \mathsf{V}_{2q}$ are the outputs of random function and XOR construction respectively, where $\mathsf{V}_1, \ldots, \mathsf{V}_{2q} \leftarrow_{\text{wor}} \{0,1\}^n$. Let $\mathbf{P_1}$ and $\mathbf{P_2}$ denote the output distributions of $\mathsf{X} := (\mathsf{X}_1, \ldots, \mathsf{X}_q)$ and $\mathsf{U}' := (\mathsf{U}'_1, \ldots, \mathsf{U}'_q)$ respectively. Thus,

$$\mathbf{Adv}^{\text{prf}}_{\mathsf{XOR}}(\mathcal{A}) \leq d_{\text{TV}}(\mathbf{P_1}, \mathbf{P_2}).$$

Now, we note that X_i's cannot take 0^n and hence it is natural to consider the q-tuple of random variables $\mathsf{U}_1, \ldots, \mathsf{U}_q \leftarrow_{\$} [N]^* := \{0,1\}^n \setminus \{0^n\}$. Let us denote by $\mathbf{P_0}$ the probability distribution of $\mathsf{U}_1, \ldots, \mathsf{U}_q$. By simple algebra, we have $d_{\text{TV}}(\mathbf{P_0}, \mathbf{P_2}) \leq q/2^n$. Also, using triangle inequality[4], we have

$$\mathbf{Adv}^{\text{prf}}_{\mathsf{XOR}}(\mathcal{A}) \leq d_{\text{TV}}(\mathbf{P_0}, \mathbf{P_1}) + q/2^n.$$

At this point, the χ^2 method (i.e., Theorem 1) gives an upper bound on $d_{\text{TV}}(\mathbf{P_0}, \mathbf{P_1})$. The rest of the proof is devoted to show $d_{\text{TV}}(\mathbf{P_0}, \mathbf{P_1}) \leq \frac{0.5q + 3\sqrt{q}}{2^n}$.

For every non-zero $x_1, \ldots, x_i$, we clearly have $\mathbf{P}_{\mathbf{0}|x^{i-1}}(x_i) = 1/(N-1)$. For simplicity, let us denote by $Y_{i,x}$ the conditional probability $\mathbf{P}_{\mathbf{1}|x^{i-1}}(x)$ which is also a function over x^{i-1}. When x^{i-1} is chosen following the distribution of X^{i-1}, we denote $Y_{i,x}$ as $\mathsf{Y}_{i,x}$. From the definition of χ^2 function corresponding to $(\mathsf{V}_1, \ldots, \mathsf{V}_q)$ and $(\mathsf{U}_1, \ldots, \mathsf{U}_q)$, we have

$$\chi^2(x^{i-1}) = \sum_{x \neq 0^n} (N-1)(Y_{i,x} - \frac{1}{N-1})^2. \tag{7}$$

Now, we give a brief description of the rest but critical part of the proof where the authors provided an upper bound on $\mathbf{Ex}[\chi^2(\mathsf{X}^{i-1})]$. We keep the authors' flow (suppressing some calculation which will be denoted as $***$) and wordings. However, we change some of their notations in order to make them consistent with our notation. Authors complete the proof as described below.

`We now expand` $Y_{i,x}$ `into a more expressive and convenient formula to work with.` $***$ `Let` $\mathsf{S} = \{\mathsf{V}_1, \mathsf{V}_2, \ldots, \mathsf{V}_{2i-2}\}$`. Let` $\mathsf{D}_{i,x}$ `be the number of pairs` $(u, u \oplus x)$ `such that both` u `and` $u \oplus x$ `belongs to` S`. Note that` S `and` $\mathsf{D}_{i,x}$ `are both random variables, and in fact functions of the random variables` $\mathsf{V}_1, \mathsf{V}_2, \ldots, \mathsf{V}_{2i-2}$`.` $***$ `Hence,`

$$\mathsf{Y}_{i,x} = \frac{N - 4(i-1) + \mathsf{D}_{i,x}}{(N-2i+1)(N-2i)}. \tag{8}$$

[4] Triangle inequality of total variation metric can be easily shown from the triangle inequality in real numbers.

$$* * *$$

$$\left(\mathsf{Y}_{i,x} - \frac{1}{N-1}\right)^2 \leq \frac{3(\mathsf{D}_{i,x} - 4(i-1)^2/N)^2 + 18}{N^4}.$$

From Eq. 7,

$$\mathbf{Ex}[\chi^2(\mathsf{X}^{i-1})] \leq \sum_{x \neq 0^n} N \cdot \mathbf{Ex}\left[\left(\mathsf{Y}_{i,x} - \frac{1}{N-1}\right)^2\right] \tag{9}$$

$$\leq \sum_{x \neq 0^n} \frac{18}{N^3} + \frac{3}{N^3} \cdot \mathbf{Ex}\left[\left(\mathsf{D}_{i,x} - \frac{4(i-1)^2}{N}\right)^2\right] \tag{10}$$

In the last formula, it is helpful to think of each $\mathsf{D}_{i,x}$ as a function of $\mathsf{V}_1, \mathsf{V}_2, \ldots, \mathsf{V}_{2i-2}$, and the expectation is taken over the choices of $\mathsf{V}_1, \mathsf{V}_2, \ldots, \mathsf{V}_{2i-2}$ sampled uniformly without replacement from $\{0,1\}^n$. We will show that[5] for any $x \in \{0,1\}^n \setminus \{0^n\}$,

$$\mathbf{Ex}\left[\left(\mathsf{D}_{i,x} - \frac{4(i-1)^2}{N}\right)^2\right] \leq \frac{4(i-1)^2}{N} \tag{11}$$

and thus

$$\mathbf{Ex}[\chi^2(\mathsf{X}^{i-1})] \leq \frac{18}{N^2} + \frac{12(i-1)^2}{N^3}.$$

Summing up, from χ^2-method

$$d_{\mathrm{TV}}(\mathbf{P_0}, \mathbf{P_1}) \leq \left(\frac{1}{2}\sum_{i=1}^{q} \mathbf{Ex}[\chi^2(\mathsf{X}^{i-1})]\right)^{\frac{1}{2}}$$
$$\leq \frac{3\sqrt{q} + .5q}{N}. \qquad \square$$

4.1 Flaw in the Above Proof and its Repair Under an Assumption

Let us revisit Eq. 8. Let us fix distinct $v_1, \ldots, v_{2i-2}$ and define the set $\mathcal{S} = \{v_1, \ldots, v_{2i-2}\}$. Let $D_{i,x}$ denote the number of pairs $(u, u \oplus x)$ such that both u and $u \oplus x$ belong to $\mathcal{S}$. Let $x_1 = v_1 \oplus v_2, \ldots, x_{i-1} = v_{2i-3} \oplus v_{2i-2}$. Now, it is easy to see that

$$\mathbf{P}(\mathsf{X}_i = x | \mathsf{V}_1 = v_1, \ldots, \mathsf{V}_{2i-2} = v_{2i-2}) = \frac{N - 4(i-1) + D_{i,x}}{(N-2i+1)(N-2i)} \tag{12}$$

which appeared in the right hand side of Eq. 8. Whereas the left hand side of the equation is $\mathbf{P}(\mathsf{X}_i = x | \mathsf{X}_1 = x_1, \ldots, \mathsf{X}_{i-1} = x_{i-1})$. Note that in general,

$$\mathbf{P}(\mathsf{X}_i = x | \mathsf{V}_1 = v_1, \ldots, \mathsf{V}_{2i-2} = v_{2i-2}) = \mathbf{P}(\mathsf{X}_i = x | \mathsf{X}_1 = x_1, \ldots, \mathsf{X}_{i-1} = x_{i-1}) \tag{13}$$

[5] Which has been shown later in the proof given by Dai et al. In this paper we don't provide details on this claim and so we skip this proof here.

does not hold for every $v_1, \ldots, v_{2i-2}$. Hence Eq. 8 is incorrect.

After observing this flaw in the proof, let us see how we can fix it. If we can prove Eq. 10 in some other way, we can still continue with the rest of the proof. This can be proved if we can prove a variant of the Eq. 8 as follows:

$$\sum_x \mathbf{Ex}[(\mathsf{Y}_{i,x} - c)^2] = \sum_x \mathbf{Ex}\left[\left(\frac{N - 4(i-1) + \mathsf{D}_{i,x}}{(N-2i+1)(N-2i)} - c\right)^2\right],$$

where $c = 1/(N-1)$. In other words,

$$\sum_x \mathbf{Ex}[(\mathbf{P}(\mathsf{X}_i = x | \mathsf{X}^{i-1}) - c)^2] = \sum_x \mathbf{Ex}[(\mathbf{P}(\mathsf{X}_i = x | \mathsf{V}^{2i-2}) - c)^2].$$

The above equation is equivalent to

$$\underset{\mathsf{V}^{2i-2}}{\mathbf{Ex}}[\sum_{x \in [N]^*} \mathbf{P}(\mathsf{X}_i = x | \mathsf{V}^{2i-2})^2] = \underset{\mathsf{X}^{i-1}}{\mathbf{Ex}}[\sum_{x \in [N]^*} \mathbf{P}(\mathsf{X}_i = x | \mathsf{X}^{i-1})^2] \tag{14}$$

We will show in the subsequent section that (14) is not actually true. This in fact shows that the Eq. 13 is actually an inequality for certain choices of v_i's. However, the proof can still survive if we have the following weaker statement which we place here as an assumption.

Assumption 1

$$\underset{\mathsf{V}^{2i-2}}{\mathbf{Ex}}[\sum_{x \in [N]^*} \mathbf{P}(\mathsf{X}_i = x | \mathsf{V}^{2i-2})^2] \geq \underset{\mathsf{X}^{i-1}}{\mathbf{Ex}}[\sum_{x \in [N]^*} \mathbf{P}(\mathsf{X}_i = x | \mathsf{X}^{i-1})^2] \tag{15}$$

Under this assumption, we can justify the inequalities appearing in Eq. 9 and Eq. 10. Hence, we can complete the proof of PRF-security of XOR construction under the above assumption. In the following section, we study the assumption for $i = 2$ and $i = 3$.

5 On the Correctness of Assumption 1

In this section, we study the correctness of our assumption stated in the previous section for some small choices of i, namely for $i = 2$ and 3.

Theorem 4.

$$\underset{\mathsf{X}_1}{\mathbf{Ex}}[\sum_{x_2 \in [N]^*} \mathbf{P}(\mathsf{X}_2 = x_2 | \mathsf{X}_1)^2] = \underset{\mathsf{V}^2}{\mathbf{Ex}}[\sum_{x_2 \in [N]^*} \mathbf{P}(\mathsf{X}_2 = x_2 | \mathsf{V}^2)^2]. \tag{16}$$

Proof. For notational simplicity we will write $[N]^*$ to represent the set $\{1, 2, \ldots, N-1\}$. Now, we evaluate the two sides of (16).

L.H.S. of (16): Expanding the l.h.s. of (16) we get

$$\mathop{\mathbf{Ex}}_{\mathsf{X}_1}[\sum_{x_2\in[N]^*}\mathbf{P}(\mathsf{X}_2=x_2|\mathsf{X}_1)^2]=\sum_{x_1}\sum_{x_2\in[N]^*}\frac{\mathbf{P}(\mathsf{X}_2=x_2,\mathsf{X}_1=x_1)^2}{\mathbf{P}(\mathsf{X}_1=x_1)}. \tag{17}$$

Now, it follows that

$$\begin{aligned}\mathbf{P}(\mathsf{X}_1=x_1)&=\mathbf{P}(\mathsf{V}^2=v^2|\mathsf{V}_1+\mathsf{V}_2=x_1)\\&=\frac{N}{N(N-1)}\\&=\frac{1}{N-1}.\end{aligned}$$

Next, we split the sum in (17) into the following two subcases: (*i*) Case 1.1, where we consider the condition $x_2=x_1$, and (*ii*) Case 1.2, where we consider the condition $x_2\neq x_1$. In each of the subcases, we determine the number of tuples (x_2,x_1) and probability $\mathbf{P}(\mathsf{X}_2=x_2,\mathsf{X}_1=x_1)$ of a particular tuple satisfying the conditions of the subcase.

Case 1.1: $(x_2=x_1)$.

$$|\{(x_2,x_1)|x_2=x_1\}|=N-1.$$

Next, we have

$$\begin{aligned}\mathbf{P}(\mathsf{X}_2=x_2,\mathsf{X}_1=x_1)&=\mathbf{P}(\mathsf{V}^4|\mathsf{V}_1+\mathsf{V}_2=x_1=x_2=\mathsf{V}_3+\mathsf{V}_4)\\&=\frac{N(N-2)}{N(N-1)(N-2)(N-3)}\\&=\frac{1}{(N-1)(N-3)}.\end{aligned}$$

Therefore,

$$\sum_{\substack{x_1,x_2\in[N]^*\\x_2=x_1}}\frac{\mathbf{P}(\mathsf{X}_2=x_2,\mathsf{V}^2=v^2)^2}{\mathbf{P}(\mathsf{V}^2=v^2)}=\frac{1}{(N-3)^2}. \tag{18}$$

Case 1.2: $(x_2\neq x_1)$.

$$|\{(x_2,x_1)|x_2\neq x_1\}|=(N-2)(N-3).$$

Now,

$$\begin{aligned}\mathbf{P}(\mathsf{X}_2=x_2,\mathsf{X}_1=x_1)&=\mathbf{P}(\mathsf{V}^4|\mathsf{V}_1+\mathsf{V}_2=x_1\neq x_2=\mathsf{V}_3+\mathsf{V}_4)\\&=\frac{N(N-4)}{N(N-1)(N-2)(N-3)}\end{aligned}$$

$$= \frac{N-4}{(N-1)(N-2)(N-3)}.$$

Hence,

$$\sum_{\substack{x_1,x_2\in[N]^* \\ x_2\neq x_1}} \frac{\mathbf{P}(\mathsf{X}_2 = x_2, \mathsf{X}_1 = x_1)^2}{\mathbf{P}(\mathsf{X}_1 = x_1)} = \frac{(N-4)^2}{(N-2)(N-3)^2}. \tag{19}$$

From (18) and (19) we get

$$\mathop{\mathbf{Ex}}_{\mathsf{X}_1}[\sum_{x_2\in[N]^*} \mathbf{P}(\mathsf{X}_2 = x_2|\mathsf{X}_1)^2] = \frac{1}{(N-3)^2} + \frac{(N-4)^2}{(N-2)(N-3)^2}. \tag{20}$$

R.H.S. of (16): Expanding the r.h.s. of (16) we get

$$\mathop{\mathbf{Ex}}_{\mathsf{V}^2}[\sum_{x_2\in[N]^*} \mathbf{P}(\mathsf{X}_2 = x_2|\mathsf{V}^2)^2] = \sum_{v^2}\sum_{x_2\in[N]^*} \frac{\mathbf{P}(\mathsf{X}_2 = x_2, \mathsf{V}^2 = v^2)^2}{\mathbf{P}(\mathsf{V}^2 = v^2)}. \tag{21}$$

Next, it follows that

$$\mathbf{P}(\mathsf{V}^2 = v^2) = \frac{1}{N(N-1)}.$$

Similar to the l.h.s., we split the sum in the r.h.s of (21) depending on the conditions (Case 2.1), and $(i\ )\ x_2 = v_1 + v_2$ $(ii\ )\ x_2 \neq v_1 + v_2$ (Case 2.2). In each of the subcases, we determine the number of tuples (x_2, v^2) and probability $\mathbf{P}(\mathsf{X}_2 = x_2, \mathsf{V}^2 = v^2)$ of a particular tuple satisfying the conditions of the subcase.

Case 2.1:$(x_2 = v_1 + v_2)$.

$$|\{(x_2, v^2)|x_2 = v_1 + v_2\}| = N(N-1).$$

$$\begin{aligned}\mathbf{P}(\mathsf{X}_2 = x_2, \mathsf{V}^2 = v^2) &= \mathbf{P}(\mathsf{V}^4 = v^4|\mathsf{V}_1 + \mathsf{V}_2 = \mathsf{V}_3 + \mathsf{V}_4 = x_2) \\ &= \frac{(N-2)}{N(N-1)(N-2)(N-3)} \\ &= \frac{1}{N(N-1)(N-3)}.\end{aligned}$$

So,

$$\sum_{v^2}\sum_{\substack{x_2\in[N]^* \\ x_2=v_1+v_2}} \frac{\mathbf{P}(\mathsf{X}_2 = x_2, \mathsf{V}^2 = v^2)^2}{\mathbf{P}(\mathsf{V}^2 = v^2)} = \frac{1}{(N-3)^2}. \tag{22}$$

Case 2.2:$(x_2 \neq v_1 + v_2)$.

$$|\{(x_2, v^2)|x_2 \neq v_1 + v_2\}| = N(N-1)(N-2).$$

$$\begin{aligned}\mathbf{P}(\mathsf{X}_2 = x_2, \mathsf{V}^2 = v^2) &= \mathbf{P}(\mathsf{V}^4 = v^4 | \mathsf{V}_1 + \mathsf{V}_2 \neq x_2 = \mathsf{V}_3 + \mathsf{V}_4) \\ &= \frac{N(N-4)}{N(N-1)(N-2)(N-3)} \\ &= \frac{N-4}{(N-1)(N-2)(N-3)}.\end{aligned}$$

Therefore,

$$\sum_{v^2} \sum_{\substack{x_2 \in [N]^* \\ x_2 \neq v_1 + v_2}} \frac{\mathbf{P}(\mathsf{X}_2 = x_2, \mathsf{V}^2 = v^2)^2}{\mathbf{P}(\mathsf{V}^2 = v^2)} = \frac{(N-4)^2}{(N-2)(N-3)^2}. \tag{23}$$

From (22) and (23) we get

$$\underset{\mathsf{V}^2}{\mathbf{Ex}}[\sum_{x_2 \in [N]^*} \mathbf{P}(\mathsf{X}_2 = x_2 | \mathsf{V}^2)^2] = \frac{1}{(N-3)^2} + \frac{(N-4)^2}{(N-2)(N-3)^2}. \tag{24}$$

The theorem follows by comparing (20) and (24). □

So we have shown that our assumption is valid for $i = 2$ (in fact these are equal). Now we show that the assumption is still valid for $i = 3$. Here we have strict inequality.

Theorem 5.

$$\underset{\mathsf{X}^2}{\mathbf{Ex}}[\sum_{x_3 \in [N]^*} \mathbf{P}(\mathsf{X}_3 = x_3 | \mathsf{X}^2)^2] < \underset{\mathsf{V}^4}{\mathbf{Ex}}[\sum_{x_3 \in [N]^*} \mathbf{P}(\mathsf{X}_3 = x_3 | \mathsf{V}^4)^2]. \tag{25}$$

Proof.As in the proof of the previous theorem, we consider the two sides of (25) separately. However, in this proof, the calculations are more involved. In order to help the reader follow the steps of the proof, we show the proof structure in Figure 1. In the figure, the root node corresponds to (25), and its two children correspond to calculation of the two sides of (25). In the remaining nodes, we show the conditions corresponding to the subcases.
L.H.S. of (25): Expanding the l.h.s. of (25) we get

$$\underset{\mathsf{X}^2}{\mathbf{Ex}}[\sum_{x_3 \in [N]^*} \mathbf{P}(\mathsf{X}_3 = x_3 | \mathsf{X}^2)^2] = \sum_{x^2} \sum_{x_3 \in [N]^*} \frac{\mathbf{P}(\mathsf{X}_3 = x_3, \mathsf{X}^2 = x^2)^2}{\mathbf{P}(\mathsf{X}^2 = x^2)}. \tag{26}$$

We split the outer sum in the r.h.s. of (26) depending on the conditions (i) $x_1 = x_2$ (Case 1.1), and (ii) $x_1 \neq x_2$ (Case 1.2). For each of these subcases and subcases of these subcases, we determine the number of tuples (x_3, x^2) and the probability $\mathbf{P}(\mathsf{X}_3 = x_3, \mathsf{X}^2 = x^2)$ of a particular tuple satisfying the conditions of the subcase.

Case 1.1:$(x_1 = x_2)$.

In this case, we have

$$\begin{aligned}\mathbf{P}(\mathsf{X}^2 = x^2) &= \mathbf{P}(\mathsf{V}^4 = v^4 | \mathsf{V}_1 + \mathsf{V}_2 = x_1 = x_2 = \mathsf{V}_3 + \mathsf{V}_4)\\ &= \frac{N(N-2)}{N(N-1)(N-2)(N-3)}\\ &= \frac{1}{(N-1)(N-3)}.\end{aligned}$$

Nex, we consider the following two subcases of this case: (i) Case 1.1.1, where we consider the condition $x_3 = x_1$, and (ii) Case 1.1.2, where we consider the condition $x_3 \neq x_1$.

Case 1.1.1:$(x_3 = x_1 = x_2)$.

$$|\{(x_3, x_1, x_2) | x_3 = x_1 = x_2\}| = N - 1.$$

$$\begin{aligned}\mathbf{P}(\mathsf{X}_3 = x_3, \mathsf{X}_1 = x_1, \mathsf{X}_2 = x_2) &= \mathbf{P}(\mathsf{V}^6 = v^6 | \mathsf{V}_1 + \mathsf{V}_2 = \mathsf{V}_3 + \mathsf{V}_4 = \mathsf{V}_5 + \mathsf{V}_6 = x_3)\\ &= \frac{N(N-2)(N-4)}{N(N-1)(N-2)(N-3)(N-4)(N-5)}\\ &= \frac{1}{(N-1)(N-3)(N-5)}.\end{aligned}$$

Case 1.1.2:$(x_3 \neq x_1 = x_2)$

$$|\{(x_3, x^2) | x_3 \neq x_1 = x_2\}| = (N-1)(N-2).$$

$$\mathbf{P}(\mathsf{X}_3 = x_3, \mathsf{X}^2 = x^2) = \mathbf{P}(\mathsf{V}^6 = v^6 | x_1 = \mathsf{V}_1 + \mathsf{V}_2 = \mathsf{V}_3 + \mathsf{V}_4 = x_2 \neq x_3 = \mathsf{V}_5 + \mathsf{V}_6). \tag{27}$$

In order to calculate the probability on the r.h.s. of (27), we first observe that for fixed x_1, number of possible choices for $(v_1, v_2) = N$. Since $v_3, v_4 \notin \{v_1, v_2\}$, so, number of possible choices for $(v_3, v_4) = N - 2$. Further, we also require $v_5, v_6 \notin \{v_1, v_2, v_3, v_4\}$. This is equivalent to the condition that $v_5 \notin S = \{v_1, v_2, v_3, v_4\} \cup \{v_1 + x_3, v_2 + x_3, v_3 + x_3, v_4 + x_3\}$.Therefore, in order to determine the number of possible choices of (v_5, v_6) we need to calculate the cardinality of the set S.

Here, we observe that in order to calculate the cardinality of S, it is enough to determine $|\{v_1\} \cap \{v_1 + x_3, v_2 + x_3, v_3 + x_3, v_4 + x_3\}|$. It is clear that $v_1 \notin \{v_1 + x_3, v_2 + x_3\}$. Hence, we are left with the following two subcases: Case 1.1.2.a and Case 1.1.2.b.

- `Case 1.1.2.a:`$(v_1 \in \{v_3 + x_3, v_4 + x_3\})$. We have the following two possibilities.

1. $v_1 = v_3 + x_3$: In this case,it also follows that $v_2 = v_4 + x_3$. So, $|S| = 4$.
2. $v_1 = v_4 + x_3$: Similar to the above subcase. Hence, $|S| = 4$.

- `Case 1.1.2.b:`($v_1 \notin \{v_3 + x_3, v_4 + x_3\}$). In this case, $|S| = 8$.

Therefore, out of total $N(N-2)$ choices of v^4, for $2N$ choices , we have that $v_1 \in \{v_3+x_3, v_4+x_3\}$. For these $2N$ choices of v^4, number of possible choices of $(v_5, v_6) = N-4$. For the remaining $N(N-4)$ choices of v^4, number of possible choices of $(v_5, v_6) = N-8$. Hence,

$$\begin{aligned}\mathbf{P}(\mathsf{X}_3 = x_3, \mathsf{X}^2 = x^2) &= \frac{2N(N-4) + N(N-4)(N-8)}{N(N-1)(N-2)(N-3)(N-4)(N-5)} \\ &= \frac{N-6}{(N-1)(N-2)(N-3)(N-5)}.\end{aligned}$$

Summing up the cases 1.1.1 and 1.1.2, we get

$$\begin{aligned}\sum_{\substack{x^2 \\ x_1=x_2}} \sum_{x_3 \in [N]^*} \frac{\mathbf{P}(\mathsf{X}_3 = x_3, \mathsf{X}^2 = x^2)^2}{\mathbf{P}(\mathsf{X}^2 = x^2)} =& \frac{(N-1)^2(N-3)}{((N-1)(N-3)(N-5))^2} \\ &+ \frac{(N-1)^2(N-2)(N-3)(N-6)^2}{((N-1)(N-2)(N-3)(N-5))^2} \\ =& \frac{1}{(N-3)(N-5)^2} + \\ & \frac{(N-6)^2}{(N-2)(N-3)(N-5)^2}. \qquad (28)\end{aligned}$$

Case 1.2:($x_1 \neq x_2$).

$$\mathbf{P}(\mathsf{X}^2 = x^2) = \mathbf{P}(\mathsf{V}^4 = v^4 | \mathsf{V}_1 + \mathsf{V}_2 = x_1 \neq x_2 = \mathsf{V}_3 + \mathsf{V}_4).$$

Here, we have that for fixed x_1 the number of possible choices of $(v_1, v_2) = N$. Now, $v_3 \notin \{v_1, v_2\} \cup \{v_1 + x_2, v_2 + x_2\}$. Also, $\{v_1, v_2\} \cap \{v_1 + x_2, v_2 + x_2\} = \emptyset$. Therefore, the number of possible choices of (v_3, v_4) is $N-4$. So,

$$\mathbf{P}(\mathsf{X}^2 = x^2) = \frac{N-4}{(N-1)(N-2)(N-3)}.$$

Next, we consider the following three subcases depending on (i) $x_3 = x_1$ (Case 1.2.1), (ii) $x_3 = x_2$ (Case 1.2.2), and (iii) $x_3 \notin \{x_1, x_2\}$ (Case 1.2.3).
Case 1.2.1:($x_3 = x_1$)

$$|\{(x_3, x^2) | x_3 = x_1 \neq x_2\}| = (N-1)(N-2).$$

By following the same argument as in Case 1.1.2, we get

$$\mathbf{P}(\mathsf{X}_3 = x_3, \mathsf{X}^2 = x^2) = \frac{N-6}{(N-1)(N-2)(N-3)(N-5)}.$$

Case 1.2.2:($x_3 = x_2$) This subcase is same as the previous case, i.e., the Case 1.2.1.
Case 1.2.3:($x_3 \notin \{x_1, x_2\}$) Here, we have the following subcases to consider depending on (i) $x_3 = x_1 + x_2$ (Case 1.2.3.a) and (ii) $x_3 \neq x_1 + x_2$ (Case 1.2.3.b).
`Case 1.2.3.a:`($x_3 = x_1 + x_2$).

$$|\{(x_3, x^2)|x_3, x_1, x_2 \text{ unequal}, x_3 = x_1 + x_2\}| = (N-1)(N-2).$$

$$\mathbf{P}(\mathsf{X}_3 = x_3, \mathsf{X}^2 = x^2) = \mathbf{P}(\mathsf{V}^6 = v^6 | \mathsf{V}_1 + \mathsf{V}_2 = x_1, \mathsf{V}_3 + \mathsf{V}_4 = x_2, \mathsf{V}_5 + \mathsf{V}_6 = x_1 + x_2).$$

Following an analysis similar to Case 1.1.2 we obtain

$$\mathbf{P}(\mathsf{X}_3 = x_3, \mathsf{X}^2 = x^2) = \frac{N-8}{(N-1)(N-2)(N-3)(N-5)}.$$

`Case 1.2.3.b:`($x_3 \neq x_1 + x_2$).

$$|\{(x_3, x^2)|x_3, x_1, x_2 \text{ unequal}, x_3 \neq x_1 + x_2\}| = (N-1)(N-2)(N-4).$$

$$\mathbf{P}(\mathsf{X}_3 = x_3, \mathsf{X}^2 = x^2) = \mathbf{P}(\mathsf{V}^6 = v^6 | \mathsf{V}_1 + \mathsf{V}_2 = x_1, \mathsf{V}_3 + \mathsf{V}_4 = x_2, \mathsf{V}_5 + \mathsf{V}_6 \neq x_1 + x_2).$$

In this case also, we follow an analysis similar to Case 1.1.2 to obtain

$$\begin{aligned}\mathbf{P}(\mathsf{X}_3 = x_3, \mathsf{X}^2 = x^2) &= \frac{N(N-8)(N-8) + 4N(N-6)}{N(N-1)(N-2)(N-3)(N-4)(N-5)} \\ &= \frac{N^2 - 12N + 40}{(N-1)(N-2)(N-3)(N-4)(N-5)}.\end{aligned}$$

By adding the cases 1.2.1, 1.2.2, 1.2.3.a, and 1.2.3.b, we obtain

$$\begin{aligned}\sum_{\substack{x^2 \\ x_1 \neq x_2}} \sum_{x_3 \in [N]^*} \frac{\mathbf{P}(\mathsf{X}_3 = x_3, \mathsf{X}^2 = x^2)^2}{\mathbf{P}(\mathsf{X}^2 = x^2)} =& \frac{2(N-6)^2}{(N-3)(N-4)(N-5)^2} + \\ & \frac{(N-8)^2}{(N-3)(N-4)(N-5)^2} + \\ & \frac{(N^2 - 12N + 40)^2}{(N-3)(N-4)^2(N-5)^2}. \end{aligned} \tag{29}$$

Finally, by adding (28) and (29), we get the following expression for the l.h.s. of (25).

$$\begin{aligned}\mathop{\mathbf{Ex}}_{\mathsf{X}^2}[\sum_{x_3\in[N]^*}\mathbf{P}(\mathsf{X}_3=x_3|\mathsf{X}^2=x^2)^2] = \frac{1}{(N-3)(N-5)^2}+ \\ \frac{(N-6)^2}{(N-2)(N-3)(N-5)^2}+ \\ \frac{2(N-6)^2}{(N-3)(N-4)(N-5)^2}+ \\ \frac{(N-8)^2}{(N-3)(N-4)(N-5)^2}+ \\ \frac{(N^2-12N+40)^2}{(N-3)(N-4)^2(N-5)^2} \\ =\frac{\mathtt{NUM}}{\mathtt{DEN}}, \qquad (30)\end{aligned}$$

where $\mathtt{NUM} = N^5 - 22N^4 + 195N^3 - 870N^2 + 1936N - 1568$ and $\mathtt{DEN} = N^6 - 23N^5 + 217N^4 - 1073N^3 + 2926N^2 - 4160N + 2400$.

R.H.S. of (16): Next, we obtain the expression for the r.h.s. of (25). As in the case of l.h.s., we expand the sum under expectation as follows.

$$\mathop{\mathbf{Ex}}_{\mathsf{V}^4}[\sum_{x_3\in[N]^*}\mathbf{P}(\mathsf{V}_5+\mathsf{V}_6=x_3|\mathsf{V}^4)^2] = \sum_{v^4}\sum_{x_3\in[N]^*}\frac{\mathbf{P}(\mathsf{V}_5+\mathsf{V}_6=x_3|\mathsf{V}^4)^2}{\mathbf{P}(\mathsf{V}^4=v^4)} \qquad (31)$$

Clearly, we have

$$\mathbf{P}(\mathsf{V}^4=v^4) = \frac{1}{N(N-1)(N-2)(N-3)}.$$

Similar to Case 1, we split the outer sum in the r.h.s. of (31) in two subcases: (*i*) in case 2.1 we consider the condition $v_1 + v_2 = v_3 + v_4$, and (*ii*) in Case 2.2, we consider the condition $v_1 + v_2 \neq v_3 + v_4$. For each of these subcases and subcases of these subcases we determine the number of tuples (x_3, v^4) and the probability $\mathbf{P}(\mathsf{V}_5 + \mathsf{V}_6 = x_3, \mathsf{V}^4 = v^4)$ of a particular tuple satisfying the conditions of the subcase.

Case 2.1:($v_1 + v_2 = v_3 + v_4$). We further divide this subcase according to the conditions (*i*) $x_3 = v_1+v_2 = v_3+v_4$ (Case 2.1.1), and (*ii*) $x_3 \neq v_1+v_2 = v_3+v_4$ (Case 2.1.2) .
Case 2.1.1:($x_3 = v_1 + v_2 = v_3 + v_4$).

$$|\{(x_3, v^4)|x_3 = v_1 + v_2 = v_3 + v_4\}| = N(N-1)(N-2).$$

$$\mathbf{P}(\mathsf{V}_5+\mathsf{V}_6=x_3,\mathsf{V}^4=v^4) = \sum_{\substack{(v_5,v_6),v_5\neq v_6,v_5+v_6=x_3\\ \{v_5,v_6\}\cap\{v_1,v_2,v_3,v_4\}=\emptyset}}(\mathbf{P}(\mathsf{V}_5=v_5,\mathsf{V}_6=v_6|\mathsf{V}^4=v^4)\times$$

$$\mathbf{P}(\mathsf{V}^4 = v^4))$$

$$= \sum_{\substack{(v_5,v_6),v_5\neq v_6,v_5+v_6=x_3\\ \{v_5,v_6\}\cap\{v_1,v_2,v_3,v_4\}=\emptyset}} \frac{\mathbf{P}(\mathsf{V}_5 = v_5, \mathsf{V}_6 = v_6|\mathsf{V}^4 = v^4)}{N(N-1)(N-2)(N-3)}.$$

Clearly, we have

$$|\{(v_5, v_6)|v_5 \neq v_6, v_5 + v_6 = x_3, \{v_5, v_6\} \cap \{v_1, v_2, v_3, v_4\} = \emptyset\}| = N - 4.$$

And for fixed (v_5, v_6) and v^4,

$$\mathbf{P}((\mathsf{V}_5 = v_5, \mathsf{V}_6 = v_6)|\mathsf{V}^4 = v^4) = \frac{1}{(N-4)(N-5)}.$$

So,

$$\mathbf{P}(\mathsf{V}_5 + \mathsf{V}_6 = x_3, \mathsf{V}^4 = v^4) = \frac{1}{N(N-1)(N-2)(N-3)(N-5)}.$$

Case 2.1.2:$(x_3 \neq v_1 + v_2 = v_3 + v_4)$. By following an argument similar to the Case 1.1.2 we arrive at the following two subcases depending on the conditions (i) $v_1 \in \{v_3 + x_3, v_4 + x_3\}$ (Case 2.1.2.a), and (ii) $v_1 \notin \{v_3 + x_3, v_4 + x_3\}$ (Case 2.1.2.b).

`Case 2.1.2.a:`$(v_1 \in \{v_3 + x_3, v_4 + x_3\})$.

$$|\{(x_3, v^4)|x_3 \neq v_1 + v_2, x_3 \in \{v_1 + v_3, v_1 + v_4\}\}| = 2N(N-1)(N-2).$$

The number of possible choices of the pair (v_5, v_6) is $N - 4$. Therefore,

$$\mathbf{P}(\mathsf{V}_5 + \mathsf{V}_6 = x_3, \mathsf{V}^4 = v^4) = \frac{1}{N(N-1)(N-2)(N-3)(N-5)}.$$

`Case 2.1.2.b:`$(v_1 \notin \{v_3 + x_3, v_4 + x_3\})$.

$$|\{(x_3, v^4)|x_3 \notin \{v_1 + v_2, v_1 + v_3, v_1 + v_4\}\}| = N(N-1)(N-2)(N-4).$$

Also, the number of possible choices of the pair (v_5, v_6) is $N - 8$. So, in this case we have

$$\mathbf{P}(\mathsf{V}_5 + \mathsf{V}_6 = x_3, \mathsf{V}^4 = v^4) = \frac{(N-8)}{N(N-1)(N-2)(N-3)(N-4)(N-5)}.$$

So, summing up the cases 2.1.1, 2.1.2.a and 2.1.2.b, we get

$$\sum_{\substack{v^4\\ v_1+v_2=v_3+v_4}} \sum_{x_3\in[N]^*} \frac{\mathbf{P}(\mathsf{V}_5 + \mathsf{V}_6 = x_3, \mathsf{V}^4 = v^4)^2}{\mathbf{P}(\mathsf{V}^4 = v^4)} = \frac{1}{(N-3)(N-5)^2} + \frac{2}{(N-3)(N-5)^2} + \frac{(N-8)^2}{(N-3)(N-4)(N-5)^2}. \quad (32)$$

Case 2.2:($v_1 + v_2 \neq v_3 + v_4$) We split this case according to the conditions (i) $x_3 \in \{v_1 + v_2, v_3 + v_4\}$ (Case 2.2.1), (ii) $x_3 = v_1 + v_2 + v_3 + v_4$ (Case 2.2.2), and (iii) $x_3 \notin \{v_1 + v_2, v_3 + v_4, v_1 + v_2 + v_3 + v_4\}$ (Case 2.2.3).

Case 2.2.1:($x_3 \in \{v_1 + v_2, v_3 + v_4\}$)

$$|\{(x_3, v^4)|x_3 \in \{v_1 + v_2, v_3 + v_4\}\}| = 2N(N-1)(N-2)(N-4).$$

In this case, possible number of choices for the pair (v_5, v_6) is $N - 6$. So, we have

$$\mathbf{P}(\mathsf{V}_5 + \mathsf{V}_6 = x_3, \mathsf{V}^4 = v^4) = \frac{N-6}{N(N-1)(N-2)(N-3)(N-4)(N-5)}.$$

Case 2.2.2:($x_3 = v_1 + v_2 + v_3 + v_4$)

$$|\{(x_3, v^4)|x_3 = v_1 + v_2 + v_3 + v_4\}\}| = N(N-1)(N-2)(N-4).$$

The number of possible choices for the pair (v_5, v_6) is $N - 8$. Hence,

$$\mathbf{P}(\mathsf{V}_5 + \mathsf{V}_6 = x_3, \mathsf{V}^4 = v^4) = \frac{N-8}{N(N-1)(N-2)(N-3)(N-4)(N-5)}.$$

Case 2.2.3:($x_3 \notin \{v_1 + v_2, v_3 + v_4, v_1 + v_2 + v_3 + v_4\}$) We further divide this case into the following two subcases depending on (i) $v_1 \in \{v_3 + x_3, v_4 + x_3\}$ (Case 2.2.3.a), and (ii) $v_1 \notin \{v_3 + x_3, v_4 + x_3\}$ (Case 2.2.3.b).

`Case 2.2.3.a:`($v_1 \in \{v_3 + x_3, v_4 + x_3\}$).

$$|\{(x_3, v^4)|x_3 \in \{v_3 + v_1, v_4 + v_1\}\}\}| = 4N(N-1)(N-2)(N-4).$$

Possible number of choices for the pair (v_4, v_5) is $N - 6$. Therefore,

$$\mathbf{P}(\mathsf{V}_5 + \mathsf{V}_6 = x_3, \mathsf{V}^4 = v^4) = \frac{N-6}{N(N-1)(N-2)(N-3)(N-5)}.$$

`Case 2.2.3.b:`($v_1 \notin \{v_3 + x_3, v_4 + x_3\}$).

$$|\{(x_3, v^4)|x_3 \notin \{v_3 + v_1, v_4 + v_1\}\}\}| = N(N-1)(N-2)(N-4)(N-8).$$

The number of possible choices of the pair (v_5, v_6) is $N - 8$. So, in this case, we have

$$\mathbf{P}(\mathsf{V}_5 + \mathsf{V}_6 = x_3, \mathsf{V}^4 = v^4) = \frac{N-8}{N(N-1)(N-2)(N-3)(N-4)(N-5)}.$$

So, summing up the subcases 2.2.1, 2.2.2, 2.2.3.a, and 2.2.3.b, we get

$$\sum_{\substack{v^4 \\ v_1+v_2\neq v_3+v_4}} \sum_{x_3\in[N]^*} \frac{\mathbf{P}(\mathsf{V}_5+\mathsf{V}_6=x_3,\mathsf{V}^4=v^4)^2}{\mathbf{P}(\mathsf{V}^4=v^4)} = \frac{2(N-6)^2}{(N-3)(N-4)(N-5)^2}+ \frac{(N-8)^2}{(N-3)(N-4)(N-5)^2}+ \frac{4(N-6)^2}{(N-3)(N-4)(N-5)^2}+ \frac{(N-8)^3}{(N-3)(N-4)(N-5)^2}. \tag{33}$$

Finally, adding (32) and (33) we get

$$\mathop{\mathbf{Ex}}_{\mathsf{V}^4}[\sum_{x_3\in[N]^*} \mathbf{P}(\mathsf{V}_5+\mathsf{V}_6=x_3|\mathsf{V}^4)^2] = \frac{(N^2-11N+36)}{(N-3)(N-4)(N-5)}. \tag{34}$$

The theorem follows by comparing (30) and (34) (difference between the two being $\frac{16(N^2-8N+8)}{(N-2)(N-3)(N-4)^2(N-5)^2}$ which is > 0 for $N \geq 8$). □

6 Concluding Remarks

Following the publication ([BN18]) of this paper, the authors of [DHT17a] have corrected the proof of Theorem 3 in [DHT17b]. Their fix of Theorem 3 is simpler by use of Jensen's ineequality.

Acknowledgement We thank the reviewers of the journal Cryptography and Communications for their comments and suggestions which have improved the quality of our manuscript.

References

BI99. M. Bellare and R. Impagliazzo, *A tool for obtaining tighter security analyses of pseudorandom function based constructions, with applications to PRP to PRF conversion*, IACR Cryptology ePrint Archive **1999**, 24 (1999).

BKR98. M. Bellare, T. Krovetz and P. Rogaway, Luby-Rackoff backwards: Increasing security by making block ciphers non-invertible, pages 266–280, Springer, 1998.

BN18. S. Bhattacharya and M. Nandi, *A note on the chi-square method: A tool for proving cryptographic security*, Cryptography and Communications **10**(5), 935–957 (2018).

BR02. J. Black and P. Rogaway, A Block-Cipher Mode of Operation for Parallelizable Message Authentication, in *EUROCRYPT 2002*, volume 2332 of *LNCS*, pages 384–397, Springer, 2002.

CLP14. B. Cogliati, R. Lampe and J. Patarin, The Indistinguishability of the XOR of k Permutations, in *Fast Software Encryption - 21st International Workshop, FSE 2014, London, UK, March 3-5, 2014. Revised Selected Papers*, edited by C. Cid and C. Rechberger, volume 8540 of *Lecture Notes in Computer Science*, pages 285–302, Springer, 2014.

CS16. B. Cogliati and Y. Seurin, EWCDM: An Efficient, Beyond-Birthday Secure, Nonce-Misuse Resistant MAC, in *CRYPTO 2016, Proceedings, Part I*, pages 121–149, 2016.

CT06. T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*, Wiley-Interscience, 2006.

DHT17a. W. Dai, V. T. Hoang and S. Tessaro, Information-theoretic Indistinguishability via the Chi-squared Method, Cryptology ePrint Archive, Report 2017/537, 2017, `http://eprint.iacr.org/2017/537`.

DHT17b. W. Dai, V. T. Hoang and S. Tessaro, *Information-theoretic Indistinguishability via the Chi-squared Method*, IACR Cryptology ePrint Archive **2017**, 537 (2017).

GG15. S. Gilboa and S. Gueron, *Distinguishing a truncated random permutation from a random function*, IACR Cryptology ePrint Archive **2015**, 773 (2015).

GG16. S. Gilboa and S. Gueron, *The Advantage of Truncated Permutations*, CoRR **abs/1610.02518** (2016).

GGM17. S. Gilboa, S. Gueron and B. Morris, *How Many Queries are Needed to Distinguish a Truncated Random Permutation from a Random Function?*, Journal of Cryptology (2017).

GLL17. S. Gueron, A. Langley and Y. Lindell, *AES-GCM-SIV: Specification and Analysis*, IACR Cryptology ePrint Archive **2017**, 168 (2017).

GS02. A. L. Gibbs and F. E. Su, *On Choosing and Bounding Probability Metrics*, International Statistical Review **70**(3), 419–435 (2002).

HWKS98. C. Hall, D. Wagner, J. Kelsey and B. Schneier, *Building PRFs from PRPs*, pages 370–389, Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.

IMPS17. T. Iwata, K. Minematsu, T. Peyrin and Y. Seurin, *ZMAC: A Fast Tweakable Block Cipher Mode for Highly Secure Message Authentication*, IACR Cryptology ePrint Archive **2017**, 535 (2017).

Iwa06. T. Iwata, New Blockcipher Modes of Operation with Beyond the Birthday Bound Security, in *Fast Software Encryption, 13th International Workshop, FSE 2006, Graz, Austria, March 15-17, 2006, Revised Selected Papers*, edited by M. J. B. Robshaw, volume 4047 of *Lecture Notes in Computer Science*, pages 310–327, Springer, 2006.

KL51. S. Kullback and R. A. Leibler, *On Information and Sufficiency*, Ann. Math. Statist. **22**(1), 79–86 (1951).

Luc00. S. Lucks, The Sum of PRPs Is a Secure PRF, in *EUROCRYPT 2000*, volume 1807 of *LNCS*, pages 470–484, Springer, 2000.

LV87. F. Liese and I. Vajda, *Convex Statistical Distances*, Teubner, Leipzig, 1987.

MN17. B. Mennink and S. Neves, Encrypted Davies-Meyer and its dual: Towards optimal security using Mirror theory, Cryptology ePrint Archive, Report 2017/xxx, to be published in CRYPTO 2017, 2017, `http://eprint.iacr.org/2017/537`.

Pat08a. J. Patarin, The "Coefficients H" Technique, in *Selected Areas in Cryptography, 2008*, volume 5381 of *LNCS*, pages 328–345, Springer, 2008.

Pat08b. J. Patarin, A Proof of Security in $O(2^n)$ for the Xor of Two Random Permutations, in *ICITS 2008*, volume 5155 of *LNCS*, pages 232–248, Springer, 2008.

Pat10. J. Patarin, Introduction to Mirror Theory: Analysis of Systems of Linear Equalities and Linear Non Equalities for Cryptography., Cryptology ePrint Archive, Report 2017/287, 2010, `http://eprint.iacr.org/2010/287`.

Rei12. R.-D. Reiss, *Approximate distributions of order statistics: with applications to nonparametric statistics*, Springer Science & Business Media, 2012.

Sli16. A. Slivkins, Lecture Notes CMSC 858G: Bandits, Experts and Games (Lecture 3), September 2016, `http://www.cs.umd.edu/~slivkins/CMSC858G-fall16/Lecture3.pdf`.

Sta78. A. J. Stam, *Distance between sampling with and without replacement*, Statistica Neerlandica **32**(2), 81–91 (1978).

Appendix A Proof of the χ^2 method

In this section we provide proof of Theorem 1, which is the heart of the χ^2 method. The proof is based on Lemma 1, Lemma 2, and Theorem 6. Along the way we also briefly mention some (relevant) facts of KL divergence and χ^2 distance.

Kullback-Leibler Divergence. *Kullback-Leibler divergence* (KL divergence) or *relative entropy* between $\mathbf{P_0}$ to $\mathbf{P_1}$ is defined as

$$d_{\mathrm{KL}}(\mathbf{P_0}, \mathbf{P_1}) = \sum_{X \in \Omega} \mathbf{P_0}(X) \log \frac{\mathbf{P_0}(X)}{\mathbf{P_1}(X)}.$$

Note that the KL divergence is defined only if $\mathbf{P_0} \ll \mathbf{P_1}$ (with the convention that $0 \log \frac{0}{0} = 0$). It was first defined by Kullback and Leibler in 1951 ([KL51]) as a generalization of the entropy notion of Shannon (see [CT06]).

It can be shown that the KL divergence between any two distributions is always non-negative (known as *Gibbs' inequality*, see [CT06]). However, it is not symmetric (i.e., $d_{\mathrm{KL}}(\mathbf{P_0}, \mathbf{P_1}) \neq d_{\mathrm{KL}}(\mathbf{P_0}, \mathbf{P_1})$ in general) and does not satisfy the triangle inequality. Thus, KL divergence is not a metric.

Though not a metric, KL divergence has some useful properties. For example, the KL divergence between any two product distributions is additive over the corresponding marginals (see [CT06], [Rei12]). The KL divergence between two joint distribution can be obtained as the sum of the KL divergences of corresponding conditional distributions. This is known as the *chain rule of KL divergence*. It is one of the crucial parts of the χ^2 method. We elaborate it in more detail below.

Chain rule of KL divergence. Let $\mathbf{P}_\mathbf{0}^{\boldsymbol{q}}$ and $\mathbf{P}_\mathbf{1}^{\boldsymbol{q}}$ be two probability distributions over Ω^q. We denote $\mathbf{P}_\mathbf{0}^{\boldsymbol{i}}$ and $\mathbf{P}_\mathbf{1}^{\boldsymbol{i}}$ to represent the marginal probability distributions for first i coordinates of $\mathbf{P}_\mathbf{0}^{\boldsymbol{q}}$ and $\mathbf{P}_\mathbf{1}^{\boldsymbol{q}}$ respectively, $1 \leq i \leq q$. In other words, if $\mathsf{X} := (\mathsf{X}_1, \ldots, \mathsf{X}_q)$ and $\mathsf{Y} := (\mathsf{Y}_1, \ldots, \mathsf{Y}_q)$ are two joint random variables following the probability distributions $\mathbf{P}_\mathbf{0}^q$ and $\mathbf{P}_\mathbf{1}^q$ then $\mathbf{P}_\mathbf{0}^{\boldsymbol{i}}$ and $\mathbf{P}_\mathbf{1}^{\boldsymbol{i}}$ represent the probability distributions of X^i and Y^i respectively. We recall that $\mathbf{P}_{\mathbf{0}|x^{i-1}}(x_i)$ denotes the conditional distribution $\mathbf{P}(\mathsf{X}_i = x_i | \mathsf{X}^{i-1} = x^{i-1})$ and similarly $\mathbf{P}_{\mathbf{1}|x^{i-1}}(x_i)$. Moreover, $\mathrm{KL}(x^{i-1}) = d_{\mathrm{KL}}(\mathbf{P}_{\mathbf{0}|x^{i-1}}, \mathbf{P}_{\mathbf{1}|x^{i-1}})$. Now we state chain rule of KL divergence.

Lemma 1 (Chain rule of KL divergence (see [CT06], Theorem 2.5.3)). *Following the above notations,*

$$d_{\mathrm{KL}}(\mathbf{P}_{\mathbf{0}}^{q}, \mathbf{P}_{\mathbf{1}}^{q}) = d_{\mathrm{KL}}(\mathbf{P}_{\mathbf{0}}^{\mathbf{1}}, \mathbf{P}_{\mathbf{1}}^{\mathbf{1}}) + \sum_{i=2}^{q} \mathbf{Ex}[\mathrm{KL}(\mathsf{X}^{i-1})].$$

Proof.

$$\begin{aligned}
d_{\mathrm{KL}}(\mathbf{P}_{\mathbf{0}}^{\boldsymbol{q}}, \mathbf{P}_{\mathbf{1}}^{\boldsymbol{q}}) &= \sum_{x^q \in \Omega^q} \mathbf{P}_{\mathbf{0}}^{\boldsymbol{q}}(x^q) \log\left(\frac{\mathbf{P}_{\mathbf{0}}^{\boldsymbol{q}}(x^q)}{\mathbf{P}_{\mathbf{1}}^{\boldsymbol{q}}(x^q)}\right) \\
&= \sum_{x^q \in \Omega^q} \mathbf{P}_{\mathbf{0}}^{\boldsymbol{q}}(x^q) \log\left(\frac{\prod_{i=1}^{q} \mathbf{P}_{\mathbf{0}|x^{i-1}}(x_i)}{\prod_{i=1}^{q} \mathbf{P}_{\mathbf{1}|x^{i-1}}(x_i)}\right) \\
&= \sum_{x^q \in \Omega^q} \mathbf{P}_{\mathbf{0}}^{\boldsymbol{q}}(x^q) \sum_{i=1}^{q} \log\left(\frac{\mathbf{P}_{\mathbf{0}|x^{i-1}}(x_i)}{\mathbf{P}_{\mathbf{1}|x^{i-1}}(x_i)}\right) \\
&= \sum_{i=1}^{q} \sum_{x^q \in \Omega^q} \mathbf{P}_{\mathbf{0}}^{\boldsymbol{q}}(x^q) \log\left(\frac{\mathbf{P}_{\mathbf{0}|x^{i-1}}(x_i)}{\mathbf{P}_{\mathbf{1}|x^{i-1}}(x_i)}\right) \\
&= \sum_{i=1}^{q} \sum_{x^i \in \Omega^i} \mathbf{P}_{\mathbf{0}}^{\boldsymbol{i}}(x^i) \log\left(\frac{\mathbf{P}_{\mathbf{0}|x^{i-1}}(x_i)}{\mathbf{P}_{\mathbf{1}|x^{i-1}}(x_i)}\right) \\
&= \sum_{i=1}^{q} \sum_{x^i \in \Omega^i} \mathbf{P}_{\mathbf{0}}^{\boldsymbol{i-1}}(x^{i-1}) \mathbf{P}_{\mathbf{0}|x^{i-1}}(x_i) \log\left(\frac{\mathbf{P}_{\mathbf{0}|x^{i-1}}(x_i)}{\mathbf{P}_{\mathbf{1}|x^{i-1}}(x_i)}\right) \\
&= \sum_{i=1}^{q} \sum_{x^{i-1} \in \Omega^{i-1}} \mathbf{P}_{\mathbf{0}}^{\boldsymbol{i-1}}(x^{i-1}) \sum_{X_i} \mathbf{P}_{\mathbf{0}|x^{i-1}}(x_i) \log\left(\frac{\mathbf{P}_{\mathbf{0}|x^{i-1}}(x_i)}{\mathbf{P}_{\mathbf{1}|x^{i-1}}(x_i)}\right) \\
&= \sum_{i=1}^{q} \sum_{x^{i-1} \in \Omega^{i-1}} \mathbf{P}_{\mathbf{0}}^{\boldsymbol{i-1}}(x^{i-1}) \mathrm{KL}(x^{i-1}) \\
&= \sum_{i=1}^{q} \mathbf{Ex}[\mathrm{KL}(\mathsf{X}^{i-1})] \qquad \square
\end{aligned}$$

The next inequality due to Pinsker (see [CT06]) gives an upper bound on the total variation distance between two distributions in terms of their KL divergence.

Theorem 6 (Pinsker's Inequality). *For every probability functions* $\mathbf{P_0}, \mathbf{P_1}$,

$$d_{\mathrm{TV}}(\mathbf{P_0}, \mathbf{P_1}) \leq \sqrt{\frac{1}{2} d_{\mathrm{KL}}(\mathbf{P_0}, \mathbf{P_1})}.$$

Proof.We follow the steps of [Sli16]. Let $\Omega' = \{x \in \Omega | \mathbf{P_0}(x) \geq \mathbf{P_1}(x)\}$. Also, let $p_i = \sum_{x \in \Omega'} \boldsymbol{P_i}(x)$ for $i \in \{0,1\}$. So, $d_{\mathrm{TV}}(\mathbf{P_0}, \mathbf{P_1}) = p_0 - p_1$. Also, by *logsum inequality*[6], we have $d_{\mathrm{KL}}(\mathbf{P_0}, \mathbf{P_1}) \geq p_0 \log \frac{p_0}{p_1} + (1-p_0) \log \frac{(1-p_0)}{(1-p_1)}$. Therefore,

$$\begin{aligned} d_{\mathrm{KL}}(\mathbf{P_0}, \mathbf{P_1}) &\geq p_0 \log \frac{p_0}{p_1} + (1-p_0) \log \frac{(1-p_0)}{(1-p_1)} \\ &= \int_{p_1}^{p_0} \left(\frac{p_0}{x} - \frac{(1-p_0)}{(1-x)} \right) dx \\ &= \int_{p_1}^{p_0} \frac{p_0 - x}{x(1-x)} dx \\ &\geq 2(p_0 - p_1)^2 = 2 d_{\mathrm{TV}}(\mathbf{P_0}, \mathbf{P_1})^2, \quad (\text{since } x(1-x) \leq \frac{1}{4}).\square \end{aligned}$$

χ^2 DISTANCE. χ^2 distance has its origin in mathematical statistics dating back to Pearson (see [LV87] for some history). The χ^2 distance between $\mathbf{P_0}$ and $\mathbf{P_1}$, with $\mathbf{P_0} \ll \mathbf{P_1}$, is defined as

$$d_{\chi^2}(\mathbf{P_0}, \mathbf{P_1}) := \sum_{x \in \Omega} \frac{(\mathbf{P_0}(x) - \mathbf{P_1}(x))^2}{\mathbf{P_1}(x)}.$$

It can be seen that χ^2 distance is not symmetric. Therefore, it is not a metric. However, like KL-divergence, χ^2 distance between product distributions can be bounded in terms of the χ^2 distances between their marginals (see [Rei12]). The following lemma shows that KL-divergence between two distributions can be upper bounded by their χ^2-distance. The first inequality can also be found in earlier works (see [GS02] for this and many other relations among various distances used in Statistics).

Lemma 2. $d_{\mathrm{KL}}(\mathbf{P_0}, \mathbf{P_1}) \leq \log(1 + d_{\chi^2}(\mathbf{P_0}, \mathbf{P_1})) \leq d_{\chi^2}(\mathbf{P_0}, \mathbf{P_1})$.

Proof. By the definition of χ^2-distance we have

$$\begin{aligned} \log(1 + d_{\chi^2}(\mathbf{P_0}, \mathbf{P_1})) &= \log \left(\sum_{x \in \Omega} \mathbf{P_0}(x) \frac{\mathbf{P_0}(x)}{\mathbf{P_1}(x)} \right) \\ &= \log \left(\mathbf{Ex} \left[\frac{\mathbf{P_0}(x)}{\mathbf{P_1}(x)} \right] \right) \\ &\geq \mathbf{Ex} \left[\log \left(\frac{\mathbf{P_0}(x)}{\mathbf{P_1}(x)} \right) \right] \text{ by Jensen's inequality} \\ &= \sum_{x \in \Omega} \mathbf{P_0}(x) \log \left(\frac{\mathbf{P_0}(x)}{\mathbf{P_1}(x)} \right) \end{aligned}$$

[6] Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be nonnegative numbers. We denote the sum $\sum_i a_i$ and $\sum_i b_i$ by a and b respectively. The log sum inequality states that $\sum_{i=1}^n a_i \log \frac{a_i}{b_i} \geq a \log \frac{a}{b}$.

$$= d_{\mathrm{KL}}(\mathbf{P_0}, \mathbf{P_1})$$

The last inequality follows by observing that $d_{\chi^2}(\mathbf{P_0}, \mathbf{P_1})) \geq 0$ and $\log(1+t) \leq t$ for $t \geq 0$. □

A.1 Proof of Theorem 1

We are now ready to show the upper bound on $d_{\mathrm{TV}}(\mathbf{P}^{\boldsymbol{q}}_{\mathbf{0}}, \mathbf{P}^{\boldsymbol{q}}_{\mathbf{1}})$ in terms of expected value of χ^2-distance between the conditional distributions $\mathbf{P}_{\mathbf{0}|x^{i-1}}$ and $\mathbf{P}_{\mathbf{1}|x^{i-1}}$. We state and prove the χ^2 method, i.e. Theorem 1.

Proof of Theorem 1. The proof follows directly from Pinsker's inequality (Theorem 6), chain rule of KL divergence (Lemma 1), and Lemma 2. More precisely, we have

$$\begin{aligned} d_{\mathrm{TV}}(\mathbf{P}^{\boldsymbol{q}}_{\mathbf{0}}, \mathbf{P}^{\boldsymbol{q}}_{\mathbf{1}}) &\leq \left(\frac{d_{\mathrm{KL}}(\mathbf{P}^{\boldsymbol{q}}_{\mathbf{0}}, \mathbf{P}^{\boldsymbol{q}}_{\mathbf{1}})}{2}\right)^{\frac{1}{2}} \text{ by Theorem 6} \\ &= \left(\frac{1}{2}\sum_{i=1}^{q} \mathbf{Ex}[\mathrm{KL}(\mathsf{X}^{i-1})]\right)^{\frac{1}{2}} \text{ by Lemma 1} \\ &\leq \left(\frac{1}{2}\sum_{i=1}^{q} \mathbf{Ex}[\chi^2(\mathsf{X}^{i-1})]\right)^{\frac{1}{2}} \text{ by Lemma 2} \qquad \square \end{aligned}$$

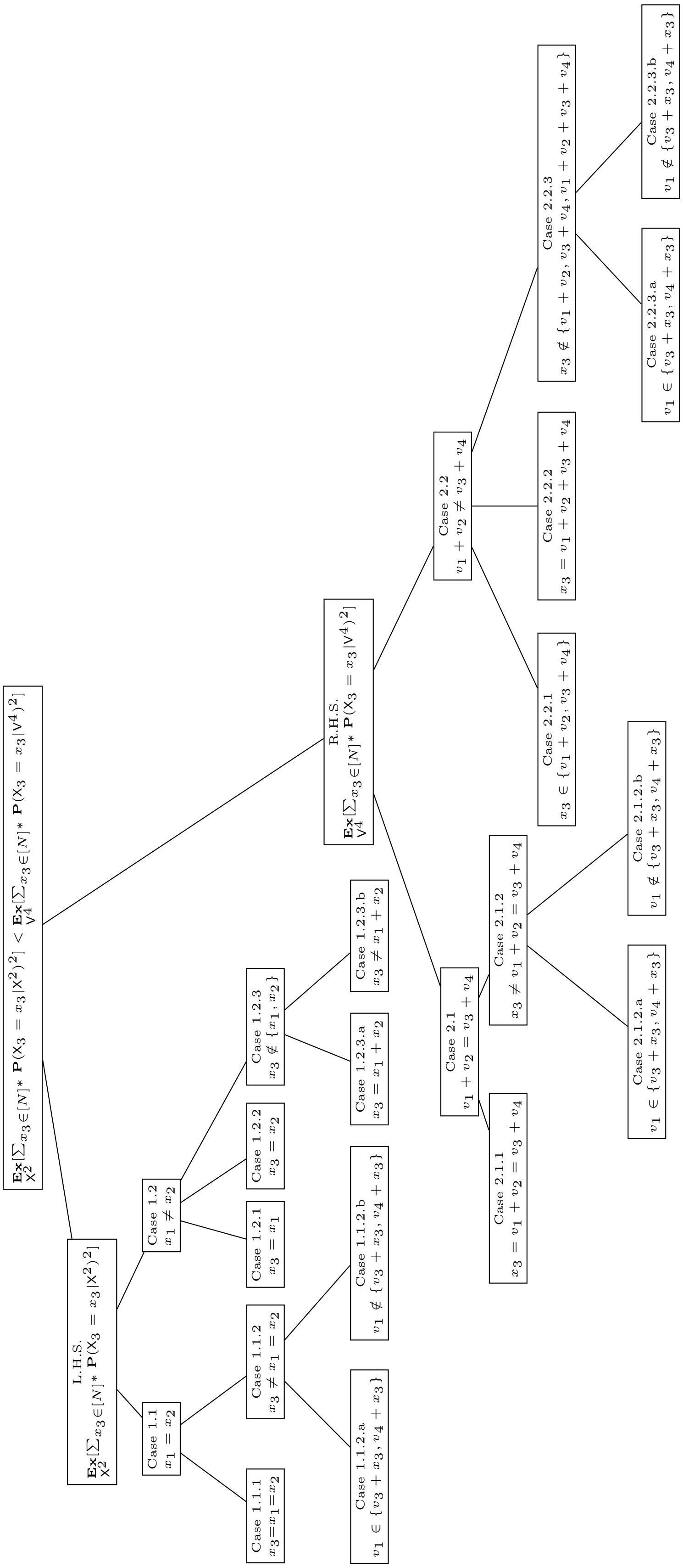

Fig. 1: Proof structure of Theorem 5