On the Security of KOS

Benjamin E. Diamond

Coinbase

benediamond@gmail.com

Abstract

We present a proof of security of the original random oblivious transfer extension protocol of Keller, Orsini, and Scholl (CRYPTO '15), without altering that protocol as written. Our result circumvents a recent negative result of Roy (CRYPTO '22), which shows that a key lemma in the original proof of KOS is false. Our proof leverages a new simulation strategy, and a careful analysis of the protocol's "correlation check". Our proof is asymptotic in nature, but suggests possible avenues for concrete security analysis.

1 Introduction

The oblivious transfer extension protocol of Keller, Orsini and Scholl [KOS15, Fig. 7] (henceforth "KOS") remains state-of-the-art, years after its introduction; it attains OT extension using only κ base OTs. Key to that protocol is a certain "correlation check", in which a number of extension OTs are "sacrificed" in a linear combination. This check is very difficult to analyze. In recent work, Roy [Roy22, § 4.1] disproves a key lemma [KOS15, Lem. 1], upon which that protocol's security analysis relies. This latter work invalidates the security proof [KOS15, Thm. 1], as originally written.

In a recent update to their work, Keller, Orsini and Scholl propose an adjusted variant of their protocol [KOS22, Fig. 10]; essentially, they suggest a special case of Roy's construction [Roy22]. Though the efficiency of the updated protocol seems comparable to the original, it is more complex, and uses different ideas. Indeed, we note that the analysis of [Roy22] is very theoretically involved.

It is of interest to analyze the security of [KOS15, Fig. 7] as *originally* written; this open problem is noted explicitly by Roy [Roy22, § 1.1], for example. This problem's motivation is not merely theoretical. In fact, many, currently-deployed production systems incorporate—and hence rely on the security of—[KOS15, Fig. 7], as originally written; the original version of that protocol is also simpler, and is easier to implement. Finally, it seems plausible that our proof techniques could be applied "in the reverse direction" to the work of [Roy22], and, in particular, that they could simplify that work, and yield analogous communication–computation tradeoffs in the malicious setting. We leave this latter task as a future research direction.

In this work, we prove the security of KOS's random OT extension protocol [KOS15, Fig. 7], without modifying that protocol. In doing so, we circumvent the failure of [KOS15, Lem. 1]. We briefly recall the details of that protocol. Informally, the rough idea of the correlation check of [KOS15, Fig. 7] is to control the row-vectors $(\mathbf{x}_i)_{i=0}^{l'-1}$ the receiver submits to the *correlated OT with errors* hybrid functionality $\mathcal{F}_{\mathsf{COTe}}^{\kappa,l'}$. If the receiver is honest, then each row $\mathbf{x}_i \in \mathbb{F}_2^\kappa$ is necessarily "monochromatic" (in the sense that its components are identical); if the receiver's rows $\mathbf{x}_i$ are *not* monochromatic, then the corrupt receiver may facilitate the distinguisher's learning certain bits of the sender's correlation vector $\Delta \in \mathbb{F}_2^\kappa$, by means of brute-force queries to the random oracle. The correlation check, at first glance, is quite natural. It prescribes that the parties jointly sample random elements $(\chi_i)_{i=0}^{l'-1}$ from $\mathbb{F}_{2^\kappa}$, using a coin-flipping functionality, that they subject their intermediate values—i.e., those they obtained from $\mathcal{F}_{\mathsf{COTe}}^{\kappa,l'}$—to a linear combination using these coefficients, and finally that they exchange the results. Specifically, the sender and receiver, having received $(\mathbf{q}_i)_{i=0}^{l'-1}$ and $(\mathbf{t}_i)_{i=0}^{l'-1}$, respectively, from $\mathcal{F}_{\mathsf{COTe}}^{\kappa,l'}$, compute $q := \sum_{i=0}^{l'-1} \chi_i \cdot \mathbf{q}_i$ and $t := \sum_{i=0}^{l'-1} \chi_i \cdot \mathbf{t}_i$, respectively; the (honest) receiver moreover computes $x := \sum_{i=0}^{l'-1} \chi_i \cdot x_i$, where $(x_i)_{i=0}^{l'-1}$ is its choice vector. Finally, the receiver sends x and t to the sender, who computes $q \overset{?}{=} t + x \cdot \Delta$. All multiplications here take place in the binary field $\mathbb{F}_{2^\kappa}$.

Informally, the correlation check controls whether the *individual* equations $\mathbf{q}_i \overset{?}{=} \mathbf{t}_i + x_i \cdot \Delta$ hold, for each $i \in \{0, \ldots, l'-1\}$, or—equivalently—whether the vector $(\mathbf{q}_i + \mathbf{t}_i + x_i \cdot \Delta)_{i=0}^{l'-1} \in \mathbb{F}_{2^\kappa}^{l'}$ is the zero vector. In actuality, however, the correlation check merely checks whether this latter vector resides within the *random hyperplane* in $\mathbb{F}_{2^\kappa}^{l'}$ given by the coefficients $(\chi_i)_{i=0}^{l'-1}$. The difficulty is that the corrupt receiver sees these coefficients, and the resulting hyperplane, *before* sending its combination results x and t; as a result, the receiver could conceivably act adaptively. Indeed, this is precisely the subtlety overlooked by [KOS15, Thm. 1]; we refer to [Roy22, § 4.1] for discussion. In particular, a dishonest receiver, conceivably, could choose x and t such a way that the vector $(\mathbf{q}_i + \mathbf{t}_i + x_i \cdot \Delta)_{i=0}^{l'-1}$ in question—though *nonzero*—nonetheless resides within this hyperplane, and thus causes the check to pass. (In fact, the receiver can set x and t arbitrarily, in general, without making reference to the values $(x_i)_{i=0}^{l'-1}$ and $(\mathbf{t}_i)_{i=0}^{l'-1}$.)

We briefly sketch the technical details of our proof. Our treatment of the corrupt sender is similar to that of [KOS15, Thm. 1] (though we supply certain details which were omitted from that proof). Our treatment of the corrupt receiver—the more difficult case—relies on a new simulation strategy for that case, as well as on careful analyses both of the distinguisher's resulting success conditions and of the protocol's correlation check. We first introduce the *majority rule* extraction strategy $x_i := \text{MAJ}_\kappa(\mathbf{x}_i)$; here, $\text{MAJ}_\kappa : \{0,1\}^\kappa \to \{0,1\}$ is the standard majority function on κ bits.

We then prove our simulator's security in roughly two steps. First, we characterize explicitly *just how much*—and how, precisely—the corrupt receiver must "cheat" in order to hand the distinguisher a non-negligible advantage (under our simulation strategy). We argue that it's necessary that the adversary submit a matrix $(\mathbf{x}_i)_{i=0}^{l'-1}$ for which the failure of monochromaticity is extreme, in a certain precise sense which we presently sketch. Following [KOS15], we write $\mathbf{e}_i := \mathbf{x}_i + x_i \cdot (1, \ldots, 1)$ for the "error vector" containing the minority bits of $\mathbf{x}_i$. We show that bits of the choice vector Δ may be learned by the distinguisher *only* by means of brute-force queries of the form $H(i \parallel \mathbf{t}_i + \mathbf{e}_i * \Delta)$, where, by induction, $\mathbf{e}_i$ introduces few *new* bit positions not already learned (say, fewer than $\log^2 \kappa$ of them). On the other hand, the distinguisher may successfully distinguish the real and ideal distributions only if it manages to query $H(i \parallel \mathbf{t}_i + \overline{\mathbf{e}_i} * \Delta)$, where here $\overline{\mathbf{e}_i}$ is some *majority* vector. We accordingly identify a key condition whereby a *majority* vector $\overline{\mathbf{e}_i}$ can be gradually "assembled" by means of a sequence of *minority* vectors $\mathbf{e}_i$; we say that the adversary's initial matrix is *flagrant* if this particular condition holds (see Definition 3.3).

We first argue that—*barring* the adversary's submitting a flagrant initial matrix—the distinguisher necessarily achieves a negligible advantage. Our proof involves a careful analysis of the distinguisher's queries. Roughly, we demonstrate that, in this case, throughout these queries, the distribution describing the projection of the unknown correlation vector Δ onto certain bits of *each* given majority vector $\overline{\mathbf{e}_i}$ nonetheless remains, in distinguisher's eyes, statistically close to a mixture of uniform distributions on high-dimensional subspaces. This structure theorem yields the desired indistinguishability in this case.

We finally show that, if the adversary *does* submit a flagrant initial matrix, it necessarily subsequently fails the correlation check with overwhelming probability. We show this by carefully analyzing the correlation check. We argue that—up to a uniform resampling of the random combination coefficients $(\chi_i)_{i=0}^{l'-1}$ used in the check—we may freely assume that the matrix $(\mathbf{e}_i)_{i=0}^{l'-1}$ is in reduced row-echelon form, and that, under the hypothesis of flagrancy, the reduced matrix must contain $\Omega(\frac{\kappa}{\log^2 \kappa})$ pivots. These pivots impose independently random $\mathbb{F}_2$-linear conditions on the unknown vector Δ. We show that, with overwhelming probability (over the choice of $(\chi_i)_{i=0}^{l'-1}$), the resulting $\mathbb{F}_2$-linear operator on Δ cannot even be *approximated* in rank by a field element x. This latter argument relies on a counting argument in $\mathbb{F}_2^{\kappa \times \kappa}$, and a union bound.

Informally, our analysis shows that, whenever the adversary cheats "enough" to materially help the distinguisher, even an *adaptive* choice of x will fail to help the adversary pass the check; indeed, in this case, with overwhelming probability—over the coefficients $(\chi_i)_{i=0}^{l'-1}$—there simply does not *exist* a field element $x \in \mathbb{F}_{2^\kappa}$ which makes the receiver's chance of passing non-negligible (in the hidden choice vector Δ). Essentially, we show that a matrix X over $\mathbb{F}_2$ with suitably many independently random columns is unlikely to reside "near" *any* field element x. It follows that—with overwhelming probability over $(\chi_i)_{i=0}^{l'-1}$—regardless of the adversary's choice of x, the rank of the linear map defining the correlation check is at least $\log^2 \kappa$; this implies the desired result. Our argument has a coding-theoretic flavor, and may be of independent interest.

We note that our analysis is asymptotic, and seems to leave open the protocol's *concrete* security. We discuss this matter further in Remark 3.11 below.

2 Background and Notation

We identify $\{0,1\} \cong \mathbb{F}_2$ as *sets*. We write κ and s throughout for a computational and statistical security parameter, respectively. We occasionally identify *vectors* in $\{0,1\}^\kappa \cong \mathbb{F}_2^\kappa$ with *subsets* of $\{0,\ldots,\kappa-1\}$, in the obvious way. We use the $*$ symbol to denote bitwise AND in $\mathbb{F}_2^\kappa$, and write w for Hamming weight. We use the symbol $\setminus$ to denote set subtraction. We fix a field structure on $\mathbb{F}_{2^\kappa}$—that is, an irreducible polynomial of degree κ in $\mathbb{F}_2[X]$—and identify $\mathbb{F}_{2^\kappa}$ with the $\mathbb{F}_2$-vectorspace $\mathbb{F}_2^\kappa$, by means of the basis $(1, X, \ldots, X^{\kappa-1})$. We write $\cdot$ for field multiplication. In what follows, we make extensive use of linear and affine-linear algebra over $\mathbb{F}_2$, without further comment; for this, we suggest the reference Cohn [Coh82, §5].

Following [KOS15, § 2], given an $l' \times \kappa$ matrix $\mathbf{x}$, we write $(\mathbf{x}_i)_{i=0}^{l'-1}$ for its rows. We write $\overline{\mathbf{x}_i}$ for the bitwise complement of a row-vector $\mathbf{x}_i \in \mathbb{F}_2^\kappa$, and $\overline{x_i}$ for the complement of a bit $x_i \in \mathbb{F}_2$. We write $\text{MAJ}_\kappa : \{0,1\}^\kappa \to \{0,1\}$ for the majority function on κ bits, defined specifically by $\mathbf{x}_i \mapsto w(\mathbf{x}_i) \geq \frac{\kappa}{2}$.

Given two probability distributions $\mathcal{Y}_0$ and $\mathcal{Y}_1$ on $\{0,1\}^\kappa$, the *statistical distance* between $\mathcal{Y}_0$ and $\mathcal{Y}_1$ is defined to be $\frac{1}{2} \cdot \sum_{\mathbf{y} \in \{0,1\}^\kappa} |\Pr[\mathcal{Y}_0 = \mathbf{y}] - \Pr[\mathcal{Y}_1 = \mathbf{y}]|$. We recall the definition of secure two-party computation (see e.g. Lindell [Lin17, § 6.6.2]).

2.1 Oblivious transfer

We recall background material on oblivious transfer, following [KOS15].

FUNCTIONALITY 2.1 ($\mathcal{F}^{\kappa}_{\mathsf{Rand}}$—coin-flipping functionality [KOS15, Fig. 5])**.**
The security parameter κ and players S and R are fixed.

- Upon receiving $(\texttt{rand}, i)$ from both players, $\mathcal{F}^{\kappa}_{\mathsf{Rand}}$ samples $\chi_i \leftarrow \mathbb{F}_2^\kappa$, and outputs $(\texttt{rand}, i, \chi_i)$ to both players.

FUNCTIONALITY 2.2 ($\mathcal{F}^{\kappa,l}_{\mathsf{COTe}}$—correlated OT with errors [KOS15, Fig. 2])**.**
The security parameter κ, the number l of resulting OTs, and players S and R are fixed.

- Upon receiving $(\texttt{initialize}, \Delta)$ from S, where $\Delta \in \mathbb{F}_2^\kappa$, $\mathcal{F}^{\kappa,l}_{\mathsf{COTe}}$ stores Δ.
- Upon receiving $\left(\texttt{extend}, (\mathbf{x}_i)_{i=0}^{l-1}\right)$ from R, where each $\mathbf{x}_i \in \mathbb{F}_2^\kappa$, $\mathcal{F}^{\kappa,l}_{\mathsf{COTe}}$ samples $\mathbf{t}_i \leftarrow \mathbb{F}_2^\kappa$ for each $i \in \{0,\ldots,l-1\}$, and outputs $\left(\texttt{extend}, (\mathbf{t}_i)_{i=0}^{l-1}\right)$ to R. $\mathcal{F}^{\kappa,l}_{\mathsf{COTe}}$ sets $\mathbf{q}_i := \mathbf{t}_i + \mathbf{x}_i * \Delta$ for each $i \in \{0,\ldots,l-1\}$, and outputs $\left(\texttt{extend}, (\mathbf{q}_i)_{i=0}^{l-1}\right)$ to S.

We note that $\mathcal{F}^{\kappa,l}_{\mathsf{COTe}}$ can be securely instantiated by the protocol of [KOS15, Fig. 3].

We moreover recall the *random OT* functionality:

FUNCTIONALITY 2.3 ($\mathcal{F}^{\kappa,l}_{\mathsf{ROT}}$—random OT functionality [KOS15, Fig. 6])**.**
The security parameter κ, the number l of resulting OTs, and players S and R are fixed.

- Upon receiving $\left(\texttt{extend}, (x_i)_{i=0}^{l-1}\right)$ from R, $\mathcal{F}^{\kappa,l}_{\mathsf{ROT}}$ samples $(\mathbf{v}_{i,0}, \mathbf{v}_{i,1}) \leftarrow \mathbb{F}_2^\kappa \times \mathbb{F}_2^\kappa$ for each $i \in \{0,\ldots,l-1\}$. $\mathcal{F}^{\kappa,l}_{\mathsf{ROT}}$ outputs $\left(\texttt{extend}, (\mathbf{v}_{i,0}, \mathbf{v}_{i,1})_{i=0}^{l-1}\right)$ to S and $\left(\texttt{extend}, (\mathbf{v}_{i,x_i})_{i=0}^{l-1}\right)$ to R.

For self-containedness, we finally recall the full protocol for $\mathcal{F}^{\kappa,l}_{\mathsf{ROT}}$, as in [KOS15, Fig. 7].

PROTOCOL 2.4 ($\Pi_{\mathsf{ROT}^{\kappa,l}}$—random OT protocol [KOS15, Fig. 7])**.**
The parameters κ and l, and players S and R, are fixed. R has input bits $(x_0,\ldots,x_{l-1})$.

- The parties write $l' := l + \kappa + s$. S samples $\Delta \leftarrow \mathbb{F}_2^\kappa$, and sends $(\texttt{intialize}, \Delta)$ to $\mathcal{F}_{\mathsf{COTe}}^{\kappa,l'}$.
- R samples random bits $x_i \leftarrow \mathbb{F}_2$, for $i \in \{l, \ldots, l'-1\}$. For each $i \in \{0, \ldots, l'-1\}$, R constructs the monochromatic vector $\mathbf{x}_i := x_i \cdot (1, \ldots, 1)$. R sends $\left(\texttt{extend}, (\mathbf{x}_i)_{i=0}^{l'-1}\right)$ to $\mathcal{F}_{\mathsf{COTe}}^{\kappa,l'}$. S and R receive $\left(\texttt{extend}, (\mathbf{q}_i)_{i=0}^{l'-1}\right)$ and $\left(\texttt{extend}, (\mathbf{t}_i)_{i=0}^{l'-1}\right)$, respectively, from $\mathcal{F}_{\mathsf{COTe}}^{\kappa,l'}$.
- For each $i \in \{0, \ldots, l'-1\}$, both parties submit $(\texttt{rand}, i)$ to $\mathcal{F}_{\mathsf{Rand}}^\kappa$, and receive $(\texttt{rand}, i, \chi_i)$. R sends $x := \sum_{i=0}^{l'-1} \chi_i \cdot x_i$ and $t := \sum_{i=0}^{l'-1} \chi_i \cdot \mathbf{t}_i$ to S. S sets $q := \sum_{i=0}^{l'-1} \chi_i \cdot \mathbf{q}_i$, and checks $q \overset{?}{=} t + x \cdot \Delta$.
- For each $i \in \{0, \ldots, l-1\}$, R sets $\mathbf{v}_{i,x_i} := H\left(i \,\|\, \mathbf{t}_i\right)$, and outputs $(\mathbf{v}_{i,x_i})_{i=0}^{l-1}$. For each $i \in \{0, \ldots, l-1\}$, S sets $\mathbf{v}_{i,0} := H\left(i \,\|\, \mathbf{q}_i\right)$ and $\mathbf{v}_{i,1} := H\left(i \,\|\, \mathbf{q}_i + \Delta\right)$, and outputs $(\mathbf{v}_{i,0}, \mathbf{v}_{i,1})_{i=0}^{l-1}$.

3 Security proof

We now prove the security of Protocol 2.4 (which itself is identical to [KOS15, Fig. 7]).

Theorem 3.1. *In the $\mathcal{F}_{\mathsf{RO}}, \mathcal{F}_{\mathsf{Rand}}^\kappa, \mathcal{F}_{\mathsf{COTe}}^{\kappa,l'}$ hybrid model, Protocol 2.4 securely computes Functionality 2.3.*

Proof. We define an appropriate simulator $\mathcal{S}$.

Corrupt sender. We first handle the case in which S is corrupt. Our treatment of this case is similar to that of [KOS15, Thm. 1]. Given a real-world adversary $\mathcal{A}$ corrupting S, $\mathcal{S}$ operates in the following way.

1. As S has no input, $\mathcal{S}$ immediately receives $\left(\texttt{extend}, (\mathbf{v}_{i,0}, \mathbf{v}_{i,1})_{i=0}^{l-1}\right)$ from $\mathcal{F}_{\mathsf{ROT}}^{\kappa,l}$.
2. $\mathcal{S}$ intercepts $\mathcal{A}$'s message $(\texttt{initialize}, \Delta)$ intended for $\mathcal{F}_{\mathsf{COTe}}^{\kappa,l'}$. For each $i \in \{0, \ldots, l'-1\}$, $\mathcal{S}$ generates $\mathbf{q}_i \leftarrow \mathbb{F}_2^\kappa$ randomly. $\mathcal{S}$ moreover programs the random oracle so that, for each $i \in \{0, \ldots, l-1\}$, $H\left(i \,\|\, \mathbf{q}_i\right) := \mathbf{v}_{i,0}$ and $H\left(i \,\|\, \mathbf{q}_i + \Delta\right) := \mathbf{v}_{i,1}$. $\mathcal{S}$ finally simulates $\mathcal{F}_{\mathsf{COTe}}^{\kappa,l'}$ sending $\mathcal{A}$ $\left(\texttt{extend}, (\mathbf{q}_i)_{i=0}^{l'-1}\right)$.
3. For each $i \in \{0, \ldots, l'-1\}$, $\mathcal{S}$ samples $\chi_i \leftarrow \mathbb{F}_2^\kappa$ randomly, and simulates $\mathcal{F}_{\mathsf{Rand}}^\kappa$ sending $\mathcal{A}$ $(\texttt{rand}, i, \chi_i)$. $\mathcal{S}$ draws $x \leftarrow \mathbb{F}_2^\kappa$, computes $q := \sum_{i=0}^{l'-1} \chi_i \cdot \mathbf{q}_i$, and sets $t := q + x \cdot \Delta$. $\mathcal{S}$ sends $\mathcal{A}$ t and x, as if from R.

The perfection of this simulation is self-evident, except perhaps for the distribution of x. For self-containedness, we present a full proof of the relevant lemma, whose proof is omitted from [KOS15, Lem. 2].

Lemma 3.2. *Given a random $\kappa \times (\kappa + s)$ matrix X over $\mathbb{F}_2$, where $s \geq 0$,* $\Pr[\operatorname{rank}(X) = \kappa] \geq 1 - 2^{-s}$.

Proof. We incorporate arguments which are somewhat classical, but which were made explicit by Brent, Gao and Lauder [BGL03]. For each fixed value $s \geq 0$, the probability that the random matrix X's κ rows are independent is equal to the probability that each of its successive rows resides outside of the linear subspace spanned by its previous rows. This probability is given by product expression below, which we manipulate in the following way:

$$\begin{aligned}
(1 - 2^{-s-1}) \cdot \cdots \cdot (1 - 2^{-s-\kappa}) &\geq 1 - \left(2^{-s-1} + \cdots + 2^{-s-\kappa}\right) && \text{(by the inequality [BGL03, § 4].)} \\
&= 1 - 2^{-s} \cdot \left(2^{-1} + \cdots + 2^{-\kappa}\right) \\
&\geq 1 - 2^{-s}.
\end{aligned}$$

This completes the proof of the lemma. □

The second summand of the quantity $x = \sum_{i=0}^{l-1} \chi_i \cdot x_i + \sum_{i=l}^{l'-1} \chi_i \cdot x_i$ computed by the receiver can be viewed as the image of $(x_i)_{i=l}^{l'-1} \in \mathbb{F}_2^{\kappa+s}$ under the linear map $\mathbb{F}_2^{\kappa+s} \to \mathbb{F}_2^\kappa$ defined by the matrix:

$$\begin{bmatrix} | & & | \\ \chi_l & \cdots & \chi_{l'-1} \\ | & & | \end{bmatrix}.$$

The lemma implies that, with probability at least $1 - 2^s$ over the choice of $(\chi_i)_{i=l}^{l'-1}$, the map induced by this matrix is surjective; it follows that, in the real-world distribution, with overwhelming probability, the image of the uniformly random point $(x_i)_{i=l}^{l'-1} \in \mathbb{F}_2^{\kappa+s}$ under this matrix is itself uniform in $\mathbb{F}_2^\kappa$, and so perfectly hides the first term $\sum_{i=0}^{l-1} \chi_i \cdot x_i$. This completes the treatment of the corrupt sender.

Corrupt receiver. We now handle the case in which the receiver R is corrupt. Given a real-world adversary $\mathcal{A}$ corrupting the receiver R, $\mathcal{S}$ operates as follows.

1. $\mathcal{S}$ simulates the existence of $\mathcal{F}_{\mathsf{COTe}}^{\kappa,l'}$, including S's role. $\mathcal{S}$ begins by sampling $\Delta \leftarrow \mathbb{F}_2^\kappa$, as S would.

2. Upon intercepting $\mathcal{A}$'s message $\left(\texttt{extend}, (\mathbf{x}_i)_{i=0}^{l'-1}\right)$ intended for $\mathcal{F}_{\mathsf{COTe}}^{\kappa,l'}$, $\mathcal{S}$ proceeds in the following way. For each $i \in \{0, \ldots, l-1\}$, $\mathcal{S}$ assigns $x_i := \text{MAJ}_\kappa(\mathbf{x}_i)$. $\mathcal{S}$ submits the choice vector $\left(\texttt{extend}, (x_i)_{i=0}^{l-1}\right)$ to $\mathcal{F}_{\mathsf{ROT}}^{\kappa,l}$, and receives $\left(\texttt{extend}, (\mathbf{v}_{i,x_i})_{i=0}^{l-1}\right)$ from $\mathcal{F}_{\mathsf{ROT}}^{\kappa,l}$.

3. $\mathcal{S}$ randomly samples $\mathbf{t}_i \leftarrow \mathbb{F}_2^\kappa$ for each $i \in \{0, \ldots l'-1\}$. For each $i \in \{0, \ldots l'-1\}$, $\mathcal{S}$ writes $\mathbf{q}_i := \mathbf{t}_i + \Delta * \mathbf{x}_i$. $\mathcal{S}$ programs the random oracle so that, for each $i \in \{0, \ldots l-1\}$, $H\left(i \parallel \mathbf{q}_i + x_i \cdot \Delta\right) := \mathbf{v}_{i,x_i}$. $\mathcal{S}$ finally returns $\left(\texttt{extend}, (\mathbf{t}_i)_{i=0}^{l'-1}\right)$ to $\mathcal{A}$, as if from $\mathcal{F}_{\mathsf{COTe}}^{\kappa,l'}$.

4. For each $i \in \{0, \ldots, l'-1\}$, $\mathcal{S}$ samples $\chi_i \leftarrow \mathbb{F}_2^\kappa$ randomly, and simulates $\mathcal{F}_{\mathsf{Rand}}^\kappa$ sending $\mathcal{A}$ $(\texttt{rand}, i, \chi_i)$. Upon receiving x and t from $\mathcal{A}$, $\mathcal{S}$ independently computes $q := \sum_{i=0}^{l'-1} \chi_i \cdot \mathbf{q}_i$, and runs the correlation check $q \overset{?}{=} t + x \cdot \Delta$. If the check fails, $\mathcal{S}$ submits $(\texttt{abort})$ to $\mathcal{F}_{\mathsf{ROT}}^{\kappa,l}$; otherwise, $\mathcal{S}$ proceeds, and $\mathcal{F}_{\mathsf{ROT}}^{\kappa,l}$ releases the output to the ideal honest party S.

We now claim that the resulting real and ideal distributions are computationally indistinguishable. We fix a distinguisher D attacking these distributions. Following [KOS15], for each $i \in \{0, \ldots, l'-1\}$, we abbreviate $\mathbf{e}_i := \mathbf{x}_i + x_i \cdot (1, \ldots, 1)$, where x_i is as extracted by $\mathcal{S}$ above; we note that necessarily $\text{MAJ}_\kappa(\mathbf{e}_i) = 0$. We observe that the strings $\mathbf{q}_i + x_i \cdot \Delta$ and $\mathbf{q}_i + \overline{x_i} \cdot \Delta$ respectively equal $\mathbf{t}_i + \mathbf{e}_i * \Delta$ and $\mathbf{t}_i + \overline{\mathbf{e}_i} * \Delta$; all of these latter quantities are known to the distinguisher *except* for Δ. If the correlation check fails, then the real and ideal distributions are identical. Conditioned on the correlation check succeeding, the simulation is perfect *except* for the fact that, for each $i \in \{0, \ldots, l-1\}$, in the real world, the relation $\mathbf{v}_{i,\overline{x_i}} = H\left(i \parallel \mathbf{q}_i + \overline{x_i} \cdot \Delta\right)$ holds, whereas, in the ideal world, $\mathbf{v}_{i,\overline{x_i}}$ is independently random.

For notational purposes, given $x \in \mathbb{F}_{2^\kappa}$, we introduce the map $F_x : \mathbb{F}_2^\kappa \to \mathbb{F}_2^\kappa$ defined by:

$$F_x : \Delta \mapsto \sum_{i=0}^{l'-1} \chi_i \cdot \mathbf{q}_i + x \cdot \Delta + t = \sum_{i=0}^{l'-1} \chi_i \cdot (\mathbf{t}_i + \mathbf{x}_i * \Delta) + x \cdot \Delta + t;$$

all quantities above are viewed as fixed constants, known to the distinguisher, *except* for the unknown Δ. Clearly, $F_x : \mathbb{F}_2^\kappa \to \mathbb{F}_2^\kappa$ is an $\mathbb{F}_2$-affine linear map. We argue that we may assume once and for all that $\mathcal{A}$ submits an "honest" value $t = \sum_{i=0}^{l'-1} \chi_i \cdot \mathbf{t}_i$. Indeed, our below arguments depend only on the dimension of the affine subspace $\{\Delta \in \mathbb{F}_2^\kappa \mid F_x(\Delta) = 0\}$, and not on its contents; $\mathcal{A}$'s use of a value $t \neq \sum_{i=0}^{l'-1} \chi_i \cdot \mathbf{t}_i$ has the effect of replacing this subspace *either* with an affine-linear subspace of $\mathbb{F}_2^\kappa$ of equal dimension *or* with the empty affine subspace (i.e., depending on whether $t + \sum_{i=0}^{l'-1} \chi_i \cdot \mathbf{t}_i$ resides within the image of $\Delta \mapsto \sum_{i=0}^{l'-1} \chi_i \cdot (\mathbf{x}_i * \Delta) + x \cdot \Delta$ or not). If the subspace is empty, then the correlation check is guaranteed to fail, and the simulation becomes trivially secure.

We thus simplify our definition, and write:

$$F_x : \Delta \mapsto \sum_{i=0}^{l'-1} \chi_i \cdot (\mathbf{x}_i * \Delta) + x \cdot \Delta.$$

In other words, we superficially ignore the affine offset, and refer to $\text{rank}(F_x)$ and $\ker(F_x)$ throughout (though, technically speaking, we allow the latter subspace to be a nonempty *affine* subspace).

We first formulate a *necessary* (though not necessarily sufficient) condition for the distinguisher's success. We identify vectors in $\mathbb{F}_2^\kappa$ with subsets of $\{0, \ldots, \kappa - 1\}$ in what follows.

Definition 3.3. Given a (truncated) initial matrix $(\mathbf{x}_i)_{i=0}^{l-1}$ submitted by $\mathcal{A}$, we write $(\mathbf{e}_i)_{i=0}^{l-1}$ for the resulting list of minority vectors, and run the following algorithm:

```
1: mark each element of the list $(\mathbf{e}_i)_{i=0}^{l-1}$ white, and initialize $\mathbf{d} := \varnothing$.
2: for $r \in \{0, \ldots, l-1\}$ do
3:     for $i \in \{0, \ldots, l-1\}$ do
4:         if $0 < |\mathbf{e}_i \setminus \mathbf{d}| < 2 \cdot \log^2 \kappa$ then
5:             mark $\mathbf{e}_i$ black.
6:             overwrite $\mathbf{d} \cup= \mathbf{e}_i$.
7:             break.
```

If, now, $|\overline{\mathbf{e}_i} \setminus \mathbf{d}| < 2 \cdot \log^2 \kappa$ holds for any $i \in \{0, \ldots, l-1\}$, then we say that $(\mathbf{x}_i)_{i=0}^{l-1}$ is *flagrant.*

Informally, the data $(\mathbf{x}_i)_{i=0}^{l-1}$ is flagrant if, by iteratively including new rows whose minority vectors $\mathbf{e}_i$ each introduce fewer than $2 \cdot \log^2 \kappa$ *new* bits, one can eventually get close to a majority vector $\overline{\mathbf{e}_i}$. We note that the vector $\mathbf{d}$ will necessarily eventually stabilize, in *at most* l iterations of the outer loop 2.

Remark 3.4. It is somewhat subtle—but true—that the vector $\mathbf{d}$ produced by Definition 3.3 depends *only* on the rows $(\mathbf{e}_i)_{i=0}^{l-1}$, and not on their ordering. As we appear not to need this fact directly, we omit its proof.

We argue that D can distinguish the real and ideal worlds *only* by means of flagrant initial matrices. Roughly, we argue that D necessarily gains little information about the bits of Δ *outside* of $\mathbf{d}$ throughout its queries; indeed, we show that, in D's view, the distribution of the projection $\left\{\overline{\mathbf{d}} * \Delta \;\middle|\; \Delta\right\}$ of Δ onto the bits outside of $\mathbf{d}$ remains statistically close to a mixture of uniform distributions on high-dimensional subspaces.

Lemma 3.5. *If $(\mathbf{x}_i)_{i=0}^{l-1}$ is* not *flagrant, then the real and ideal distributions are indistinguishable.*

Proof. We first note that we may freely assume that $\text{rank}(F_x) < \log^2 \kappa$, and prove the result only in this setting; indeed, the event in which $\text{rank}(F_x) \geq \log^2 \kappa$ *and* $\mathcal{A}$ passes the correlation check occurs with probability at most $2^{-\log^2 \kappa}$ in the random vector Δ, which is negligible. We recall that the real and ideal distributions are identical *unless* D queries $H(i \parallel \mathbf{t}_i + \overline{\mathbf{e}_i} * \Delta)$, for some $i \in \{0, \ldots, l-1\}$. On the other hand, D may learn information about Δ by means of brute-force queries of the form $\mathbf{v}_{i,x_i} \stackrel{?}{=} H\left(i \parallel \mathbf{t}_i + \mathbf{r}\right)$, where $i \in \{0, \ldots, l-1\}$ and $\mathbf{r} \in \{\mathbf{e}_i * \Delta \mid \Delta \in \ker(F_x)\}$; specifically, upon each such query, D may rule in or out (i.e., depending on whether equality holds) the candidate $\mathbf{r}$ for the value of the projection $\mathbf{e}_i * \Delta$. We argue that we may assume that D never submits a query $H\left(i \parallel \mathbf{t}_i + \mathbf{r}\right)$ for which $\mathbf{r} \neq \mathbf{e}_i * \Delta$ but nonetheless $\mathbf{v}_{i,x_i} = H\left(i \parallel \mathbf{t}_i + \mathbf{r}\right)$ holds; indeed, each such query yields a spurious equality with only negligible probability (over the random oracle's coins). In particular, we assume below that $\mathbf{v}_{i,x_i} = H\left(i \parallel \mathbf{t}_i + \mathbf{r}\right)$ implies $\mathbf{r} = \mathbf{e}_i * \Delta$.

We write $\mathbf{d}$ for the vector assembled by Definition 3.3. We write $W := \{\mathbf{d} * \Delta \mid \Delta \in \ker(F_x)\}$ for the image of the projection of $\ker(F_x)$ onto $\mathbf{d}$, and $Y := \left\{\overline{\mathbf{d}} * \Delta \;\middle|\; \Delta \in \ker(F_x)\right\}$ for $\ker(F_x)$'s projection onto $\overline{\mathbf{d}}$. For each $\mathbf{w} \in W$, we write $Y^{\mathbf{w}} := \left\{\overline{\mathbf{d}} * \Delta \;\middle|\; \mathbf{d} * \Delta = \mathbf{w} \wedge \Delta \in \ker(F_x)\right\}$; that is, $Y^{\mathbf{w}}$ is the projection onto $\overline{\mathbf{d}}$ of those $\Delta \in \ker(F_x)$ whose projection $\mathbf{d} * \Delta = \mathbf{w}$ onto $\mathbf{d}$ is suitably prescribed. Each $Y^{\mathbf{w}}$ is an affine subspace of Y. In fact, as $\mathbf{w} \in W$ varies, the corresponding sets $Y^{\mathbf{w}}$ yield a family of nonempty, disjoint—though not necessarily distinct, in that different $\mathbf{w}$ may yield identical $Y^{\mathbf{w}}$—affine subspaces of Y, of equal dimension. Indeed, each $Y^{\mathbf{w}}$ may be viewed as (the isomorphic projection onto $\overline{\mathbf{d}}$ of) the intersection in $\mathbb{F}_2^\kappa$ between $\ker(F_x)$ and the $\left|\overline{\mathbf{d}}\right|$-dimensional affine-linear subspace $\{\Delta \in \mathbb{F}_2^\kappa \mid \mathbf{d} * \Delta = \mathbf{w}\}$. By hypothesis on $\text{rank}(F_x)$, even in the extreme case in which these subspaces intersect transversely in $\mathbb{F}_2^\kappa$, we necessarily nonetheless have $\dim(Y^{\mathbf{w}}) > \left|\overline{\mathbf{d}}\right| - \log^2 \kappa$. In any case, as the intersecting affine subspaces $\{\Delta \in \mathbb{F}_2^\kappa \mid \mathbf{d} * \Delta = \mathbf{w}\}$, as $\mathbf{w} \in W$ varies, are parallel, the dimensions of the resulting intersections $Y^{\mathbf{w}} \subset Y$ are identical.

We now study D's posterior distribution $\mathcal{Y} := \left\{\overline{\mathbf{d}} * \Delta \;\middle|\; \Delta\right\}$ regarding Δ, and, in particular, how this distribution evolves as a result of D's queries to the oracle. Clearly, $\mathcal{Y}$ is supported within Y. We first note the natural conditional expansion:

$$\mathcal{Y} = \sum_{\mathbf{w} \in W} \Pr[\mathbf{d} * \Delta = \mathbf{w}] \cdot (\mathcal{Y} \mid Y^{\mathbf{w}}); \tag{1}$$

we recall once more that unequal vectors $\mathbf{w} \in W$ may nonetheless yield equal sets $Y^{\mathbf{w}}$. We now claim that, roughly, this expansion represents $\mathcal{Y}$ as a mixture of distributions which are statistically close to uniform:

Condition. For each $\mathbf{w} \in W$, the conditional $\mathcal{Y} \,|\, Y^{\mathbf{w}}$ is either empty or statistically close to uniform on $Y^{\mathbf{w}}$.

We argue that D's queries preserve the condition. We first extract the following technical claim:

Claim 3.6. *Let the distribution $\mathcal{Y}$ on Y satisfy the condition, and fix a vector $\mathbf{f}$ for which $|\mathbf{f} \setminus \mathbf{d}| \geq 2 \cdot \log^2 \kappa$. For $\mathbf{y} \leftarrow \mathcal{Y}$ sampled randomly, the probability that D outputs $\mathbf{f} * \mathbf{y}$ in polynomial time is negligible.*

Proof. We let $\mathcal{Y}$ and $\mathbf{f}$ be as in the hypothesis of the claim. We write $Z^{\mathbf{w}} := \{\mathbf{f} * \mathbf{y} \mid \mathbf{y} \in Y^{\mathbf{w}}\}$ and $\mathcal{Z} := \{\mathbf{f} * \mathbf{y} \mid \mathbf{y} \leftarrow \mathcal{Y}\}$ for the projections onto $\mathbf{f}$ of each set $Y^{\mathbf{w}}$ and of $\mathcal{Y}$, respectively. By hypothesis on $\mathbf{f}$, as each $\dim(Y^{\mathbf{w}}) > \left|\overline{\mathbf{d}}\right| - \log^2 \kappa$, we see that each $\dim(Z^{\mathbf{w}}) > \log^2 \kappa$. Moreover, for fixed $\mathbf{w}^* \in W$ and arbitrary $\mathbf{w} \in W$, the affine subspaces $Z^{\mathbf{w}^*}$ and $Z^{\mathbf{w}}$ of $\left\{\mathbf{f} * \overline{\mathbf{d}} * \Delta \;\middle|\; \Delta \in \mathbb{F}_2^{\kappa}\right\}$ are either disjoint or identical, and in any case are parallel and of equal dimension. Each $\mathcal{Z} \mid Z^{\mathbf{w}^*}$ is thus a mixture of those distributions $\left\{\mathbf{f} * \mathbf{y} \mid \mathbf{y} \leftarrow \mathcal{Y} \mid Y^{\mathbf{w}}\right\}$ for which $Z^{\mathbf{w}} = Z^{\mathbf{w}^*}$. Using our hypothesis on $\mathcal{Y}$, and the surjectivity of each such $Y^{\mathbf{w}}$'s projection onto $Z^{\mathbf{w}^*}$, we see that each element in this mixture is either empty or statistically close to uniform on $Z^{\mathbf{w}^*}$; we conclude that $\mathcal{Z} \mid Z^{\mathbf{w}^*}$ too is. We see that $\mathcal{Z}$ is a mixture of close-to-uniform distributions on superpolynomially-sized sets; as $\mathbf{f} * \mathbf{y}$ for $\mathbf{y} \leftarrow \mathcal{Y}$ random is just a sample from $\mathcal{Z}$, the conclusion follows. □

We now argue by induction that D's queries preserve the condition. We note that, initially, $\mathcal{Y} \mid Y^{\mathbf{w}}$ is exactly uniform on $Y^{\mathbf{w}}$ for each $\mathbf{w} \in W$, and that the condition clearly holds. We now consider the effect on $\mathcal{Y}$ of an arbitrary query $\mathbf{v}_{i,x_i} \overset{?}{=} H\left(i \,\|\, \mathbf{t}_i + \mathbf{r}\right)$, say, where $\mathbf{r} \in \{\mathbf{e}_i * \Delta \mid \Delta \in \ker(F_x)\}$, or, equivalently, of the information $\mathbf{r} \overset{?}{=} \mathbf{e}_i * \Delta$. We treat two separate cases, corresponding to whether $\mathbf{e}_i \overset{?}{\subset} \mathbf{d}$.

We first consider the case $\mathbf{e}_i \subset \mathbf{d}$. We write $W_{\mathbf{r}} := \{\mathbf{w} \in W \mid \mathbf{e}_i * \mathbf{w} = \mathbf{r}\}$; clearly, $W_{\mathbf{r}} \subset W$ is a nonempty affine-linear subspace. We note that the effect of the information $\mathbf{r} = \mathbf{e}_i * \Delta$ is to set $\Pr[\mathbf{d} * \Delta = \mathbf{w}] := 0$ for each $\mathbf{w} \notin W_{\mathbf{r}}$; similarly, the effect of the information $\mathbf{r} \neq \mathbf{e}_i * \Delta$ is to set $\Pr[\mathbf{d} * \Delta = \mathbf{w}] := 0$ for each $\mathbf{w} \in W_{\mathbf{r}}$. In each case, the information merely reallocates certain coefficients in the mixture expression (1); the condition is nonetheless clearly preserved.

We now treat the case $\mathbf{e}_i \not\subset \mathbf{d}$. In this case, Definition 3.3 implies that in fact $|\mathbf{e}_i \setminus \mathbf{d}| \geq 2 \cdot \log^2 \kappa$ holds. We first argue that we may assume that D's query fails, in the sense that $\mathbf{r} \neq \mathbf{e}_i * \Delta$. Indeed, by induction, we have that the condition holds on $\mathcal{Y}$. As $\overline{\mathbf{d}} * \Delta$ is distributed exactly according to $\mathcal{Y}$ (by definition), the claim, applied to the vector $\mathbf{f} := \mathbf{e}_i$, implies that the equality $\overline{\mathbf{d}} * \mathbf{r} \overset{?}{=} \mathbf{e}_i * \overline{\mathbf{d}} * \Delta$ holds with at most negligible probability; we see that, with overwhelming probability, $\overline{\mathbf{d}} * \mathbf{r} \neq \overline{\mathbf{d}} * \mathbf{e}_i * \Delta$, and hence $\mathbf{r} \neq \mathbf{e}_i * \Delta$. By analogy with the above, we write $W_{\mathbf{r}} := \{\mathbf{w} \in W \mid \mathbf{e}_i * \mathbf{w} = \mathbf{d} * \mathbf{r}\}$. For each $\mathbf{w} \notin W_{\mathbf{r}}$, the information $\mathbf{r} \neq \mathbf{e}_i * \Delta$ has no effect on $\mathcal{Y}|Y^{\mathbf{w}}$. On the other hand, for each $\mathbf{w} \in W_{\mathbf{r}}$, the information $\mathbf{r} \neq \mathbf{e}_i * \Delta$ has the effect of excluding from consideration those candidates $\overline{\mathbf{d}} * \Delta$ in $Y^{\mathbf{w}}_{\mathbf{r}} \subset Y^{\mathbf{w}}$, where we write $Y^{\mathbf{w}}_{\mathbf{r}} := \left\{\mathbf{y} \in Y^{\mathbf{w}} \;\middle|\; \mathbf{e}_i * \mathbf{y} = \overline{\mathbf{d}} * \mathbf{r}\right\}$. To show that the condition is preserved, we show that the ratio $\left|Y^{\mathbf{w}}_{\mathbf{r}}\right| / \left|Y^{\mathbf{w}}\right|$ is negligible for each $\mathbf{w} \in W_{\mathbf{r}}$. Each $Y^{\mathbf{w}}_{\mathbf{r}} \subset Y^{\mathbf{w}}$ is a linear subspace, obviously *contained* in the linear subspace $\left\{\overline{\mathbf{d}} * \Delta \;\middle|\; \mathbf{e}_i * \overline{\mathbf{d}} * \Delta = \overline{\mathbf{d}} * \mathbf{r}\right\}$ of dimension $\left|\overline{\mathbf{d}} \setminus \mathbf{e}_i\right|$. Using our above estimate on $\dim(Y^{\mathbf{w}})$, we see that:

$$\dim(Y^{\mathbf{w}}) - \dim(Y^{\mathbf{w}}_{\mathbf{r}}) > \left|\overline{\mathbf{d}}\right| - \left|\overline{\mathbf{d}} \setminus \mathbf{e}_i\right| - \log^2 \kappa = \left|\overline{\mathbf{d}} \cap \mathbf{e}_i\right| - \log^2 \kappa = |\mathbf{e}_i \setminus \mathbf{d}| - \log^2 \kappa \geq \log^2 \kappa,$$

where, in the final equality, we use our hypothesis $|\mathbf{e}_i \setminus \mathbf{d}| \geq 2 \cdot \log^2 \kappa$. We thus see that $\left|Y^{\mathbf{w}}_{\mathbf{r}}\right| / \left|Y^{\mathbf{w}}\right|$ is negligible. By this fact, and our inductive hypothesis whereby $\mathcal{Y} \mid Y^{\mathbf{w}}$ is statistically close to uniform, we conclude that $\mathcal{Y} \mid (Y^{\mathbf{w}} \setminus Y^{\mathbf{w}}_{\mathbf{r}})$ also is. This completes the argument that the condition on $\mathcal{Y}$ is preserved throughout D's queries.

We now fix an arbitrary majority vector $\overline{\mathbf{e}_i}$, for $i \in \{0, \ldots, l-1\}$, and fix a query $\mathbf{v}_{i,\overline{x_i}} \overset{?}{=} H\left(i \,\|\, \mathbf{t}_i + \mathbf{r}\right)$, where $\mathbf{r} \in \{\overline{\mathbf{e}_i} * \Delta \mid \Delta \in \ker(F_x)\}$. Finally using the hypothesis of the lemma, we note that $|\overline{\mathbf{e}_i} \setminus \mathbf{d}| \geq 2 \cdot \log^2 \kappa$. Applying Claim 3.6 once again, now with $\mathbf{f} := \overline{\mathbf{e}_i}$, we see that the equality $\overline{\mathbf{d}} * \mathbf{r} \overset{?}{=} \overline{\mathbf{e}_i} * \overline{\mathbf{d}} * \Delta$ holds with at most negligible probability, and hence that $\mathbf{r} \overset{?}{=} \overline{\mathbf{e}_i} * \Delta$ also does. This completes the proof of the lemma. □

In what follows, we thus assume that $\mathcal{A}$'s initial data $(\mathbf{x}_i)_{i=0}^{l-1}$ is flagrant. We argue that, in any such execution, it is negligibly probable that $\mathcal{A}$ will pass the correlation check. To this end, we analyze the correlation check more closely. It suffices to show that, for any flagrant initial matrix $(\mathbf{x}_i)_{i=0}^{l-1}$, it is unlikely—over the choice of the $(\chi_i)_{i=0}^{l'-1}$—that $\mathcal{A}$ will be able to find *any* $x \in \mathbb{F}_{2^\kappa}$ for which $\operatorname{rank}(F_x)$ is low.

In fact, we show that for any flagrant $(\mathbf{x}_i)_{i=0}^{l-1}$, it holds with overwhelming probability (over the choice of $(\chi_i)_{i=0}^{l'-1}$) that the minimal rank

$$\min_{x \in \mathbb{F}_{2^\kappa}} \operatorname{rank}(F_x) \tag{2}$$

is at least $\log^2 \kappa$.

We begin our study of the family of maps $\{F_x\}_{x \in \mathbb{F}_{2^\kappa}}$. It is clear that, at the cost of adding $\sum_{x_i=1} \chi_i$ to each x in the expression (2) (which has no effect), we may freely replace each $\mathbf{x}_i$ with $\mathbf{e}_i$ in the definition of F_x. We thus further rewrite F_x as follows:

$$F_x : \Delta \mapsto \sum_{i=0}^{l'-1} \chi_i \cdot (\mathbf{e}_i * \Delta) + x \cdot \Delta.$$

We moreover argue that the random variable (2) (viewed as a function of the random coefficients $(\chi_i)_{i=0}^{l'-1}$) remains identical if we replace the matrix $(\mathbf{e}_i)_{i=0}^{l'-1}$ with its reduced row-echelon form over $\mathbb{F}_2$. Indeed, any F_x may be decomposed into the $\mathbb{F}_2$-linear map $\Delta \mapsto (\mathbf{e}_i * \Delta)_{i=0}^{l'-1}$ from $\mathbb{F}_2^\kappa \to \mathbb{F}_{2^\kappa}^{l'}$, on the one hand, followed by the application of the random $\mathbb{F}_{2^\kappa}$-hyperplane given by $(\chi_i)_{i=0}^{l'-1}$, on the other (and finally by the addition of $x \cdot \Delta$). Row-reducing $(\mathbf{e}_i)_{i=0}^{l'-1}$ amounts to interposing between these first two maps a further $l' \times l'$ invertible matrix over $\mathbb{F}_{2^\kappa}$. Up to a uniform resampling of the coefficients $(\chi_i)_{i=0}^{l'-1}$, this multiplication has no effect.

We record the following claim:

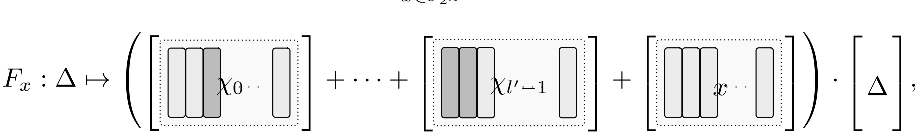
Lemma 3.7. *If $(\mathbf{x}_i)_{i=0}^{l'-1}$ is flagrant, then the $\mathbb{F}_2$-row-reduction of $(\mathbf{e}_i)_{i=0}^{l'-1}$ contains at least $\frac{\kappa}{4 \cdot \log^2 \kappa}$ pivots.*

Proof. As a matrix's number of pivots depends only on its rank, it suffices to prove the lemma after arbitrarily permuting $(\mathbf{e}_i)_{i=0}^{l'-1}$'s rows and columns. We thus first sort the rows $(\mathbf{e}_i)_{i=0}^{l'-1}$ in the order in which they are marked black by Definition 3.3 (deferring white rows). Moreover, we apply the following modification to the Gaussian elimination algorithm. By definition, each successive black row necessarily introduces a 1 to some column which thus far has lacked one. Upon each such row's treatment by the algorithm, after possibly transposing two columns, we may ensure that this 1 resides at the column being considered for a pivot, and thus becomes a pivot. This transposition preserves the invariant whereby each *further* black row introduces a 1 at some new column. Likewise, using the new pivot row to clear the pivot column also preserves this invariant. We thus conclude that there are at least as many pivots as there are black rows.

Finally, we note that there must be strictly more than $\frac{\kappa/2 - 2 \cdot \log^2 \kappa}{2 \cdot \log^2 \kappa}$ black rows in any flagrant matrix. Indeed, each row marked black may, by definition, increase the Hamming weight $w(\mathbf{d})$ only by less than $2 \cdot \log \kappa$; as each majority row satisfies $w(\overline{\mathbf{e}_i}) \geq \frac{\kappa}{2}$, the conclusion follows. □

Henceforth, we write $\widehat{\kappa}$ for the number of pivots in $(\mathbf{e}_i)_{i=0}^{l'-1}$.

We observe that each map in the family $\{F_x\}_{x \in \mathbb{F}_{2^\kappa}}$ can be written using the following matrix expression:



(3)

where the field elements x and $(\chi_i)_{i=0}^{l'-1}$ are viewed as $\mathbb{F}_2$-linear operators on $\mathbb{F}_2^\kappa$, and hence represented as $\kappa \times \kappa$ $\mathbb{F}_2$-*matrices*, and the shaded boxes indicate that certain columns have been "struck out". Indeed, we keep or strike columns of the matrices $(\chi_i)_{i=0}^{l'-1}$ according to the (row-reduced) data $(\mathbf{e}_i)_{i=0}^{l'-1}$; specifically, if $\mathbf{e}_{i,j} = 1$, we keep the j^{th} column of χ_i's matrix intact, and otherwise replace it with a column of 0s.

We resume our consideration of the expression (2), viewed as a random variable on the coefficients $(\chi_i)_{i=0}^{l'-1}$. In light of our assumption that $(\mathbf{e}_i)_{i=0}^{l'-1}$ is row-reduced, we see that each pivot in the matrix $(\mathbf{e}_i)_{i=0}^{l'-1}$ adds an *independent random column* to the matrix expression (3) (i.e., to the left-hand sum, *excluding* x). We argue that we may consider the pivot columns *alone* in our study of (2). Indeed, replacing each non-pivot column with a column of 0s—in *all* matrices within the expression (3), including that of x—can only *decrease* the rank of the resulting map F_x; we shall lower-bound this rank regardless.

We're thus left to consider the following modified expression for F_x:

$$F_x : \Delta \mapsto \left(\begin{bmatrix} X \cdots \end{bmatrix} + \begin{bmatrix} x \cdots \end{bmatrix} \right) \cdot \begin{bmatrix} \Delta \end{bmatrix}, \tag{4}$$

where the first matrix, say X, contains $\widehat{\kappa}$ *independently* random columns, with its further columns identically 0, and where the second matrix is merely the field-multiplication matrix of x, with the same set of $\widehat{\kappa}$ columns kept and the rest struck out.

We now show that, with overwhelming probability over the choice of the random sub-matrix X, we have $\min_{x \in \mathbb{F}_{2^\kappa}} \operatorname{rank}(F_x) \geq \log^2 \kappa$. We achieve this using a counting argument in $\mathbb{F}_2^{\kappa \times \kappa}$; more precisely, the argument takes place in $\mathbb{F}_2^{\kappa \times \widehat{\kappa}}$. Slightly abusing notation, we identify field elements $x \in \mathbb{F}_{2^\kappa}$ with (appropriately stricken) matrices in $\mathbb{F}_2^{\kappa \times \widehat{\kappa}}$. We note that there are exactly 2^κ distinct field elements $x \in \mathbb{F}_{2^\kappa}$, and hence at most 2^κ distinct corresponding matrices. On the other hand, for each matrix $X \in \mathbb{F}_2^{\kappa \times \widehat{\kappa}}$ for which, for some $x \in \mathbb{F}_{2^\kappa}$, $\operatorname{rank}(X + x) < \log^2 \kappa$ holds, we necessarily have that $X + x = Y$, where $Y \in \mathbb{F}_2^{\kappa \times \widehat{\kappa}}$ is of rank less than $\log^2 \kappa$. We undertake to count such matrices Y.

Lemma 3.8. *The number of matrices* $Y \in \mathbb{F}_2^{\kappa \times \widehat{\kappa}}$ *satisfying* $\operatorname{rank}(Y) < \log^2 \kappa$ *is in* $2^{\widetilde{O}(\kappa)}$.

Proof. We again refer to Brent, Gao and Lauder [BGL03] for preliminaries on subspaces over finite fields. The proof of [BGL03, Lem. 4] gives an expression for the number of matrices of rank r in $\mathbb{F}_2^{\kappa \times \widehat{\kappa}}$, which we in turn crudely upper-bound as follows:

$$\Phi_{\widehat{\kappa}}(\kappa, r) = \prod_{i=0}^{r-1} \left(2^{\widehat{\kappa}} - 2^i \right) \cdot \frac{2^{\kappa - i} - 1}{2^{i+1} - 1} \leq 2^{(\widehat{\kappa} + \kappa) \cdot r} \leq 2^{2 \cdot \kappa \cdot r}$$

Upper-bounding the number of rank-r $\kappa \times \widehat{\kappa}$ matrices by the number of rank-$\log^2 \kappa$ such matrices, for each $r < \log^2 \kappa$, we see that the total number of $\kappa \times \widehat{\kappa}$ matrices of rank *less than* $\log^2 \kappa$ is at most $\log^2 \kappa \cdot 2^{2 \cdot \kappa \cdot \log^2 \kappa}$. This quantity is clearly in $2^{\widetilde{O}(\kappa)}$, as desired. This completes the proof of the lemma. □

The set of matrices $X \in \mathbb{F}_2^{\kappa \times \widehat{\kappa}}$ for which $\min_{x \in \mathbb{F}_{2^\kappa}} \operatorname{rank}(X + x) < \log^2 \kappa$ is exactly the union, over all field elements $x \in \mathbb{F}_{2^\kappa}$, of the sets $\left\{ x + Y \,\middle|\, \operatorname{rank}(Y) < \log^2 \kappa \right\} \subset \mathbb{F}_2^{\kappa \times \widehat{\kappa}}$. In light of Lemma 3.8, we conclude that the cardinality of this union is at most $2^\kappa \cdot 2^{\widetilde{O}(\kappa)}$, which is itself in $2^{\widetilde{O}(\kappa)}$. Finally, the *total* number of $\kappa \times \widehat{\kappa}$ matrices X is obviously $2^{\kappa \cdot \widehat{\kappa}}$. The probability, over the random coefficients $(\chi_i)_{i=0}^{l'-1}$, that $\min_{x \in \mathbb{F}_{2^\kappa}} \operatorname{rank}(F_x) < \log^2 \kappa$ is thus at most $2^{\widetilde{O}(\kappa) - \kappa \cdot \widehat{\kappa}}$. From Lemma 3.7, we recall that $\widehat{\kappa} \geq \frac{\kappa}{4 \cdot \log^2 \kappa}$; we conclude that $2^{\widetilde{O}(\kappa) - \kappa \cdot \widehat{\kappa}} \leq 2^{-\Omega(\kappa)}$, which is negligible. This completes the proof of the theorem. □

We discuss a few illustrative examples.

Example 3.9. If R is honest, then the matrix of error vectors $(\mathbf{e}_i)_{i=0}^{l'-1}$ is identically zero, the matrix X above is likewise empty, and R may—by setting $x := 0$ (or really, $x := \sum_{x_i = 1} \chi_i$)—cause F_x to be the zero linear map, and pass the correlation check with probability 1.

Example 3.10. The flagrant matrix *par excellence* consists of a $\kappa \times \kappa$ identity submatrix $(\mathbf{x}_i)_{i=0}^{\kappa-1} = I_\kappa$, with $(\mathbf{x}_i)_{i=\kappa}^{l'-1}$ identically zero. If a corrupt receiver R were to manage to pass the correlation check with this matrix, then the distinguisher could trivially distinguish the real and ideal distributions with probability 1 using only $O(\kappa)$ queries to the random oracle. Indeed, in this setting, for each $i \in \{0, \ldots, \kappa - 1\}$, the equality $\mathbf{v}_{i,0} \stackrel{?}{=} H\left(i \parallel \mathbf{t}_i\right)$ would if hold and only if $\Delta_i \stackrel{?}{=} 0$ did. After making κ such queries, D would thus learn the entire choice vector Δ, and could therefore easily check the equalities $\mathbf{v}_{i,1} \stackrel{?}{=} H\left(i \parallel \mathbf{t}_i + \overline{\mathbf{x}_i} * \Delta\right)$, for each $i \in \{0, \ldots, l - 1\}$. In the real world, these equalities necessarily hold; in the ideal world, they do not.

Yet having submitted such an initial matrix $(\mathbf{x}_i)_{i=0}^{l'-1}$, $\mathcal{A}$ would face—as its matrix X above—a uniformly random $\kappa \times \kappa$ matrix with independent columns. The arguments given above show that, given such a matrix randomly sampled, it is unlikely that there exists an element $x \in \mathbb{F}_{2^\kappa}$ for which $\operatorname{rank}(X + x)$ is low.

The above examples suggest an informal way to understand the proof. The only way for the adversary to help the distinguisher is to act in such a way that the matrix $(\mathbf{e}_i)_{i=0}^{l'-1}$ of error vectors contains unequal, low-weight rows. But the more the adversary does this, the more X above becomes a "patchwork" consisting of columns selected from *different* random field elements $\chi_i \leftarrow \mathbb{F}_{2^\kappa}$. As X changes in this way, it becomes *less* likely to reside near some intact field element x.

Remark 3.11. Our analysis above is essentially asymptotic in nature, and seems to leave open the protocol's concrete security (e.g., at the value $\kappa = 256$). Indeed, the implicit function $-\Omega(\kappa)$ in the exponent, in the final step of the proof of Theorem 3.1, takes a long time to become negative. Specifically, this function—something like $2 \cdot \log\log\kappa + 2 \cdot \kappa \cdot \log^2 \kappa - \frac{\kappa^2}{4 \cdot \log^2 \kappa}$—becomes negative, for the first time and thenceforth, *only* at the value $\kappa = 1,386,267$; in this light, our proof is arguably ineffective for values κ smaller than this one.

On the other hand, our analysis contains various sources of looseness, including, most notably, our consideration of the pivot columns *alone* in the expression (4) above. Discarding the non-pivot columns in the expression (3) makes, in practice, the rank of F_x much lower. It seems conceivable that the rank of (3)—instead of that of (4)—could be directly studied, and that a tighter analysis could result. This avenue would require the consideration of matrices X whose columns are *not* independent.

There is a further means by which our analysis could be sharpened. The final step of Theorem 3.1 is information-theoretic in nature; it shows that—with (asymptotically!) overwhelming probability in $(\chi_i)_{i=0}^{l'-1}$—a field element x suitably minimizing the rank of $X + x$ *doesn't exist.* It seems possible—even at "low" values κ, for which our information-theoretic argument remains vacuous—that field elements x of the desired form could be computationally difficult to produce (over and above the question of their existence).

We leave these prospects as future research directions.

I would like to thank Marcel Keller and Peter Scholl for a discussion which led to Remark 3.11.

References

[BGL03] Richard P. Brent, Shuhong Gao, and Alan G. B. Lauder. Random Krylov spaces over finite fields. *SIAM Journal on Discrete Mathematics*, 16(2):276–287, 2003.

[Coh82] P. M. Cohn. *Algebra*, volume 1. John Wiley & Sons, second edition, 1982.

[KOS15] Marcel Keller, Emmanuela Orsini, and Peter Scholl. Actively secure OT extension with optimal overhead. In Rosario Gennaro and Matthew Robshaw, editors, *Advances in Cryptology – CRYPTO 2015*, volume 9215 of *Lecture Notes in Computer Science*, pages 724–741, Berlin, Heidelberg, 2015. Springer Berlin Heidelberg.

[KOS22] Marcel Keller, Emmanuela Orsini, and Peter Scholl. Actively secure OT extension with optimal overhead. Unpublished update, `https://eprint.iacr.org/2015/546.pdf`, September 2022.

[Lin17] Yehuda Lindell. *Tutorials on the Foundations of Cryptography: Dedicated to Oded Goldreich*, chapter How to Simulate It – A Tutorial on the Simulation Proof Technique, pages 277–346. Information Security and Cryptography. Springer International Publishing, 2017.

[Roy22] Lawrence Roy. SoftSpokenOT: Quieter OT extension from small-field silent VOLE in the minicrypt model. In Yevgeniy Dodis and Thomas Shrimpton, editors, *Advances in Cryptology – CRYPTO 2022*, volume 13507 of *Lecture Notes in Computer Science*, pages 657–687, Cham, 2022. Springer Nature Switzerland.