

More Efficient Oblivious Transfer Extensions with Security for Malicious Adversaries*

Gilad Asharov[†] Yehuda Lindell[‡] Thomas Schneider[§] Michael Zohner[‡]

November 21, 2017

Abstract

Oblivious transfer (OT) is one of the most fundamental primitives in cryptography and is widely used in protocols for secure two-party and multi-party computation. As secure computation becomes more practical, the need for practical large scale oblivious transfer protocols is becoming more evident. Oblivious transfer extensions are protocols that enable a relatively small number of “base-OTs” to be utilized to compute a very large number of OTs at low cost. In the semi-honest setting, Ishai et al. (CRYPTO 2003) presented an OT extension protocol for which the cost of each OT (beyond the base-OTs) is just a few hash function operations. In the malicious setting, Nielsen et al. (CRYPTO 2012) presented an efficient OT extension protocol for the setting of active adversaries, that is secure in the random oracle model.

In this work, we present an OT extension protocol for the setting of malicious adversaries that is more efficient and uses less communication than previous works. In addition, our protocol can be proven secure in both the random oracle model, and in the standard model with a type of correlation robustness. Given the importance of OT in many secure computation protocols, increasing the efficiency of OT extensions is another important step forward to making secure computation practical.

Note (November 21, 2017): This version includes an important fix of the protocol for the case of a corrupted sender.

Keywords: Oblivious transfer extensions, concrete efficiency, secure computation

*This work was partially supported by the European Union’s Seventh Framework Program (FP7/2007-2013) grant agreement n. 609611 (PRACTICE).

[†]School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem, Israel. Supported by the Israeli Centers of Research Excellence (I-CORE) Program (Center No. 4/11). asharov@cs.huji.ac.il

[‡]Department of Computer Science, Bar-Ilan University, Israel. Supported by the European Research Council under the European Union’s Seventh Framework Programme (FP/2007-2013) / ERC consolidators grant agreement n. 615172 (HIPS). lindell@biu.ac.il

[§]Department of Computer Science, TU Darmstadt, Darmstadt, Germany. Supported by the DFG as part of project E4 within the CRC 1119 CROSSING, by the German Federal Ministry of Education and Research (BMBF) within EC SPRIDE and CRISP, and by the Hessian LOEWE excellence initiative within CASED. {thomas.schneider,michael.zohner}@crisp-da.de

© IACR 2015. This article is a minor revision of the version that was published by Springer-Verlag at EUROCRYPT 2015.

1 Introduction

1.1 Background

Oblivious Transfer (OT), introduced by Rabin [Rab81], is a fundamental cryptographic protocol involving two parties, a sender and a receiver. In the most commonly used 1-out-of-2 version [EGL85], the sender has a pair of messages (x_0, x_1) and the receiver has a selection bit r ; at the end of the protocol the receiver learns x_r (but nothing about x_{1-r}) and the sender learns nothing at all about r . Oblivious transfer is a fundamental tool for achieving secure computation, and plays a pivotal role in the Yao protocol [Yao86] where OT is needed for every bit of input of the client, and in the GMW protocol [GMW87] where OT is needed for every AND gate in the Boolean circuit computing the function.

Protocols for secure computation provide security in the presence of adversarial behavior. A number of adversary models have been considered in the literature. The most common adversaries are: *passive* or *semi-honest adversaries* who follow the protocol specification but attempt to learn more than allowed by inspecting the protocol transcript, and *active* or *malicious adversaries* who run any arbitrary strategy in an attempt to break the protocol. In both these cases, the security of a protocol guarantees that nothing is learned by an adversary beyond its legitimate output. Another notion is that of security in the presence of *covert adversaries*; in this case the adversary may follow any arbitrary strategy, but is guaranteed to be caught with good probability if it attempts to cheat. The ultimate goal in designing efficient protocols is to construct protocols that are secure against strong (active or covert) adversaries while adding very little overhead compared to the passive variant. In our paper, we focus primarily on the case of active adversaries, but also provide a variant for covert security.

OT extensions. As we have mentioned, OT is used extensively in protocols for secure computation. In many cases, this means several millions of oblivious transfers must be run, which can become prohibitively expensive. Specifically, the state-of-the-art protocol for achieving OT with security in the presence of active adversaries of [PVW08] achieves approximately 350 random OTs per second on standard PCs. However, a million OTs at this rate would take over 45 minutes. In order to solve this problem, OT extensions [Bea96] can be used. An OT extension protocol works by running a small number of “*base-OTs*” depending on the security parameter (e.g., a few hundred) that are used as a base for obtaining many OTs via the use of cheap symmetric cryptographic operations only. This is conceptually similar to public-key encryption where instead of encrypting a large message using RSA, which would be too expensive, a hybrid encryption scheme is used such that the RSA computation is only carried out to encrypt a symmetric key, which is then used to encrypt the large message. Such an OT extension can be achieved with extraordinary efficiency; specifically, the protocol of [IKNP03] for passive adversaries requires only three hash function computations per OT (beyond the initial base-OTs). In [ALSZ13], by applying additional algorithmic and cryptographic optimizations, the cost of OT extension for passive adversaries is so low that essentially the communication is the bottleneck. To be concrete, 10,000,000 OTs on random inputs (which suffices for many applications) can be carried out in just 2.62 seconds over a LAN with four threads [ALSZ13].

For active adversaries, OT extensions are somewhat more expensive. Prior to this work, the best protocol known for OT extensions with security against active adversaries was introduced by [NNOB12]. The computational cost of the protocol is due to the number of base-OTs needed for obtaining security, the number of symmetric operations (e.g., hash function computations)

needed for every OT in the extension, and the bandwidth. Relative to the passive OT extension of [IKNP03], the run-time of [NNOB12] is approximately 4 times longer spent on the base-OTs, 1.7 times the cost for each OT in the extension, and 2.7 times the communication. Asymptotically, regarding the number of base-OTs, for security parameter κ (e.g., $\kappa = 128$), it suffices to run κ base-OTs in the passive case. In contrast, [NNOB12] require $\lceil \frac{8}{3}\kappa \rceil$ base-OTs.

Applications of OT for malicious adversaries. Most prominently, OT is heavily used in today’s most efficient protocols for *secure computation* that allow two or more parties to securely evaluate a function expressed as Boolean circuit on their private inputs. Examples include Yao’s garbled circuits-based approaches such as [LP07, LPS08, PSSW09, KSS12, FN13, SS13, LR14, HKK⁺14, FJN14] where OTs are needed for each input, or the Tiny-OT [NNOB12, LOS14] and MiniMac protocols [DZ13, DLT14] where OTs are needed for each AND gate. Additional applications include the *private set intersection* protocol of [DCW13] which is based purely on OT, and the Yao-based *zero-knowledge* protocol of [JKO13] which allows a party to prove in zero-knowledge a predicate expressed as Boolean circuit, and needs one OT per bit of the witness.

In many of the above applications, the number of oblivious transfers needed can be huge. For instance, for many applications of practical interest, the two-party and multiparty protocols of [NNOB12, LOS14, DZ13, DLT14, DCW13] can require several hundred millions of OTs, making the cost of OT the bottleneck in the protocol. Thus, improved OT extensions immediately yield faster two-party and multi-party protocols for secure computation. We give concrete examples for typical applications requiring a large number of OTs next:

Example 1.1 *The AES circuit has $\sim 10,000$ AND gates (cf. [NNOB12]) and uses $\sim 2^{19}$ OTs when evaluated with TinyOT (≥ 40 OTs (aBits) per AND gate [LOS14]).*

Example 1.2 *The PSI circuit (Sort-Compare-Shuffle) of [HEK12] has $O(bn \log n)$ AND gates and for $n = 65,536$ elements with $b = 32$ -bits the circuit has 2^{25} AND gates and takes $\sim 2^{30}$ OTs when evaluated with TinyOT.*

Example 1.3 *The malicious PSI protocol of [DCW13] needs $1.44kn$ malicious OTs, for $n = 65,536$ elements and security parameter $k = 128$, this amounts to $\sim 12 \cdot 10^6$ OTs (~ 180 OTs per element).*

1.2 Our Contributions

In this paper, we present a new protocol for OT extensions with security in the presence of malicious adversaries, which outperforms the most efficient existing protocol of [NNOB12]. We follow the insights of prior work [ALSZ13, FN13], which show that the bottleneck for efficient OT extension is the communication, and focus on decreasing the communication at the cost of slightly increased computation. Furthermore, our protocol can be instantiated with different parameters, allowing us to tradeoff communication for computation. This is of importance since when running over a LAN the computation time is more significant than when running over a WAN where the communication cost dominates. We implement and compare our protocol to the semi-honest protocol of [IKNP03] (with optimizations of [KK13, ALSZ13]) and the malicious protocol of [NNOB12] (with optimizations of [FN13]). As can be seen from the summary of our results given in Table 1, our actively secure protocol performs better than the previously fastest protocol of [NNOB12] running at under 60% the cost of the base-OTs of [NNOB12], 70% of the cost of each OT in the extension, and 55% of the communication in the local setting. Due to the lower communication, the improvement

of our protocol over [NNOB12] in the cloud setting (between US East and Europe and thus with higher latency), is even greater with approximately 45% of the time of the base-OTs and 55% of the time for each OT in the extension.

Comparing our protocol to the passive OT extension of [IKNP03], our actively secure protocol in the local (LAN) setting costs only 133% more run-time in the base-OTs, 20% more run-time for each OT in the extension, and 50% more communication. In the cloud setting, the cost for each OT in the extension is 63% more than [IKNP03] (versus 293% more for [NNOB12]). Finally, we obtain covert security at only a slightly higher cost than passive security (just 10% more for each OT in the extension in the local setting, and 30% more in the cloud setting). Our protocol reduces the number of base-OTs that are required to obtain malicious security from $\frac{8}{3}\kappa$ for [NNOB12] to $\kappa + \epsilon\rho$, where ρ is the statistical security parameter (e.g., $\rho=40$) and $\epsilon \geq 1$ is a parameter for trading between computation and communication. To be concrete, for $\kappa=128$ -bit security, our protocol reduces the number of base-OTs from 342 to 190 in the local and to 174 in the cloud setting.

| Prot. | Security | Run-Time | | Communication | |
|-------------|----------|--------------------------|---------------------------|---------------------------|-----------------------|
| | | Local | Cloud | Local | Cloud |
| [IKNP03] | passive | $0.3s+1.07\mu s \cdot t$ | $0.7s+4.24\mu s \cdot t$ | $4KB+128bit \cdot t$ | |
| This | covert | $0.6s+1.18\mu s \cdot t$ | $1.2s+5.48\mu s \cdot t$ | $21KB+166bit \cdot t$ | |
| [Lar14] | active | - | - | $42KB+106,018bit \cdot t$ | |
| [PVW08] | active | $2975.32\mu s \cdot t$ | $4597.27\mu s \cdot t$ | $0.3KB+1,024bit \cdot t$ | |
| [NNOB12] | active | $1.2s+1.82\mu s \cdot t$ | $2.9s+12.43\mu s \cdot t$ | $43KB+342bit \cdot t$ | |
| This | active | $0.7s+1.29\mu s \cdot t$ | $1.3s+6.92\mu s \cdot t$ | $24KB+191bit \cdot t$ | $22KB+175bit \cdot t$ |

Table 1: Run-time and communication for t random OT extensions with $\kappa=128$ -bit security (amortized over 2^{26} executions; [PVW08] amortized over 2^{14} executions). 1KB= 8,192bit.

In addition to being more efficient, we can prove the security of a variant of our protocol with a version of correlation robustness (where the secret value is chosen with high min-entropy, but not necessarily uniformly), and do not require a random oracle (see §3.3). In contrast, [NNOB12] is proven secure in the random oracle model. ¹

1.3 Related Work

The first efficient OT extension protocol for semi-honest adversaries was given in [IKNP03]. Improvements and optimizations to the protocol of [IKNP03] were given in [KK13, ALSZ13].

Due to its importance, a number of previous works have tackled the question of OT extensions with security for malicious/active adversaries. There exist several approaches for achieving security against active adversaries for OT extensions. All of the known constructions build on the semi-honest protocol of [IKNP03], and add *consistency checks* of different types to the OT extension protocol, to ensure that the receiver sent consistent values. (Note that in [IKNP03], the sender cannot cheat and so it is only necessary to enforce honest behavior for the receiver.)

The first actively-secure version of OT extension used a cut-and-choose technique and was already given in [IKNP03]. This cut-and-choose technique achieves a security of 2^{-n} by performing n parallel evaluations of the basic OT extension protocol.

¹It is conjectured that the [NNOB12] OT can be proven secure without a random oracle, but this has never been proven.

This was improved on by [Nie07, HIKN08], who show that active security can be achieved at a much lower cost. Their approach works in the random oracle model and ensures security against a malicious receiver by adding a low-cost check per extended OT, which uses the uncertainty of the receiver in the choice bit of the sender. As a result, a malicious receiver who wants to learn p choice bits of the sender risks being caught with probability 2^{-p} . However, this measure allows a malicious sender to learn information about the receiver’s choice bits. They prevent this attack by combining $S \in \{2, 3, 4\}$ OTs and ensuring the security of one OT by sacrificing the remaining $S - 1$ OTs. Hence, their approach adds an overhead of at least $S \geq 2$ compared to the semi-honest OT extension protocol of [IKNP03] for a reasonable number of OTs (with $S = 2$ and approximately 10^7 OTs, they achieve security except with probability 2^{-25} , cf. [Nie07]). However, the exact complexity for this approach has not been analyzed.

An alternative approach for achieving actively-secure OT extension was given in [NNOB12]. Their approach also works in the random oracle model but, instead of performing checks per extended OT as in [Nie07, HIKN08], they perform consistency checks per base-OT. Their consistency check method involves hashing the strings that are transferred in the base-OTs and is highly efficient. In their approach, they ensure the security of a base-OT by sacrificing another base-OT, which adds an overhead of factor 2. In addition, a malicious receiver is able to learn p choice bits of the sender with probability 2^{-p} . [NNOB12] shows that this leakage can be tolerated by increasing the number of base-OTs from κ to $\lceil \frac{4}{3}\kappa \rceil$. Overall, their approach increases the number of base-OTs that has to be performed by a *multiplicative factor* of $\frac{8}{3}$. The [NNOB12] protocol has been optimized and implemented on a GPU in [FN13]. We give a full description of the [NNOB12] protocol with optimizations of [FN13] in Appendix §B.

An approach for achieving actively-secure OT extension that works in the standard model has recently been introduced in [Lar14]. Their approach achieves less overhead in the base-OTs at the expense of substantially more communication during the check routine (cf. Table 1), and is therefore considerably less efficient. Nevertheless, we point out that the work of [Lar14] is of independent interest since it is based on the original correlation robustness assumption only.

Since it is the previous best, we compare our protocol to that of [NNOB12]. Our approach reduces the number of base-OTs by removing the “sacrifice” step of [NNOB12] (where one out of every 2 base-OTs are opened) but increases the workload in the consistency check routine. Indeed, we obtain an additive factor of a statistical security parameter, instead of the multiplicative increase of [NNOB12]. This can be seen as a trade-off between reducing communication through fewer base-OTs while increasing computation through more work in the consistency check routine. We empirically show that this results in a more efficient actively secure OT extension protocol, which only has 20% more time and 50% more communication than the passively secure OT extension protocol of [IKNP03] in the local setting.

The above works all consider the concrete efficiency of OT extensions. The theoretical feasibility of OT extensions was established in [Bea96], and further theoretical foundations were laid in [LZ13].

2 Preliminaries

2.1 Notation

Our protocol uses a computational (symmetric) security parameter κ and a statistical security parameter ρ . Asymptotically, this means that our protocols are secure for any adversary running

in time $\text{poly}(\kappa)$, except with probability $\mu(\kappa) + 2^{-\rho}$. (Formally, the output distribution of a real protocol execution can be distinguished from the output distribution of an ideal execution of the OT functionality with probability at most $\mu(\kappa) + 2^{-\rho}$. See [LP11] for a formal definition of secure computation with both a statistical and computational security parameter.) In our experiments we set $\kappa = 128$ and $\rho = 40$, which is considered to be secure beyond 2020².

2.2 Oblivious Transfer

Oblivious transfer (OT) was first introduced by Rabin [Rab81] as a function where a receiver receives a message, sent by a sender, with probability $1/2$, while the sender remains oblivious whether the message was received. It was later re-defined to the functionality more commonly used today by [EGL85], where a sender inputs two messages (x_0, x_1) and the receiver inputs a choice bit r and obliviously receives x_r without learning any information about x_{1-r} . Formally, the 1-out-of-2 OT functionality on n bit strings is defined as $OT_n((x_0, x_1), r) = (\lambda, x_r)$ where λ denotes the empty string and $x_0, x_1 \in \{0, 1\}^n$. In this paper we focus on the general (and most applicable) functionality, which is equivalent to m invocations of the 1-out-of-2 OT functionality on n bit strings. That is, the sender holds as input m pairs of n -bit strings (x_j^0, x_j^1) for $1 \leq j \leq m$ and the receiver holds m selection bits $\mathbf{r} = (r_1, \dots, r_m)$. The output of the receiver is $(x_1^{r_1}, \dots, x_m^{r_m})$ while the sender has no output. We denote this functionality as $m \times OT_n$. The parties are called sender P_S and receiver P_R .

Several protocols for OT based on different cryptographic assumptions and attacker models were introduced. Most notable are the passive-secure OT protocol of [NP01] and the active-secure OT protocol of [PVW08], which are among the most efficient today. However, the impossibility result of [IR88] showed that OT protocols require costly asymmetric cryptography, which greatly limits their efficiency.

2.3 OT Extension

In his seminal work, Beaver [Bea96] introduced *OT extension* protocols, which extend few costly *base-OTs* using symmetric cryptography only. While the first construction of [Bea96] was inefficient and mostly of theoretical interest, the protocol of [IKNP03] showed that OT can be extended efficiently and with very little overhead.

Recently, the passively secure OT extension protocol of [IKNP03] was improved by [KK13, ALSZ13] who showed how the communication from P_R to P_S can be reduced by a factor of two. Furthermore, [ALSZ13] implemented and optimized the protocol and demonstrated that the main bottleneck for semi-honest OT extension has shifted from computation to communication. We give the passively secure OT extension protocol of [IKNP03] with optimizations from [ALSZ13, KK13] in Protocol 1.

2.4 On the Malicious Security of [IKNP03]

The key insight to understanding how to secure OT extension against malicious adversaries is to understand that a malicious party only has very limited possibilities for an attack. In fact, the original OT extension protocol of [IKNP03] already provides security against a malicious P_S . In addition, the only attack for a malicious P_R is in Step 2a of Protocol 1, where P_R computes and

²According to the summary of cryptographic key length recommendations at <http://keylength.com>.

sends $\mathbf{u}^i = \mathbf{t}^i \oplus G(\mathbf{k}_i^1) \oplus \mathbf{r}$ (cf. [IKNP03]). A malicious P_R could choose a different \mathbf{r} for each \mathbf{u}^i (for $1 \leq i \leq \ell$), and thereby extract P_S 's choice bits \mathbf{s} . Hence, malicious security can be obtained if P_R can be forced to use the same choice bits \mathbf{r} in all messages $\mathbf{u}^1, \dots, \mathbf{u}^\ell$.

PROTOCOL 1 (Passive-secure OT extension protocol of [IKNP03])

- **Input of P_S :** m pairs (x_j^0, x_j^1) of n -bit strings, $1 \leq j \leq m$.
- **Input of P_R :** m selection bits $\mathbf{r} = (r_1, \dots, r_m)$.
- **Common Input:** Symmetric security parameter κ and $\ell = \kappa$.
- **Oracles and cryptographic primitives:** The parties use an ideal $\ell \times OT_\kappa$ functionality, pseudorandom generator $G : \{0, 1\}^\kappa \rightarrow \{0, 1\}^m$ and a correlation robust-function $H : [m] \times \{0, 1\}^\ell \rightarrow \{0, 1\}^n$ (see §3.3).

1. *Initial OT Phase:*

- (a) P_S initializes a random vector $\mathbf{s} = (s_1, \dots, s_\ell) \in \{0, 1\}^\ell$ and P_R chooses ℓ pairs of seeds $\mathbf{k}_i^0, \mathbf{k}_i^1$ each of size κ .
- (b) The parties invoke the $\ell \times OT_\kappa$ -functionality, where P_S acts as the *receiver* with input \mathbf{s} and P_R acts as the *sender* with inputs $(\mathbf{k}_i^0, \mathbf{k}_i^1)$ for every $1 \leq i \leq \ell$.

For every $1 \leq i \leq \ell$, let $\mathbf{t}^i = G(\mathbf{k}_i^0)$. Let $T = [\mathbf{t}^1 | \dots | \mathbf{t}^\ell]$ denote the $m \times \ell$ bit matrix where its i th column is \mathbf{t}^i for $1 \leq i \leq \ell$. Let \mathbf{t}_j denote the j th row of T for $1 \leq j \leq m$.

2. *OT Extension Phase:*

- (a) P_R computes $\mathbf{t}^i = G(\mathbf{k}_i^0)$ and $\mathbf{u}^i = \mathbf{t}^i \oplus G(\mathbf{k}_i^1) \oplus \mathbf{r}$, and sends \mathbf{u}^i to P_S for every $1 \leq i \leq \ell$.
- (b) For every $1 \leq i \leq \ell$, P_S defines $\mathbf{q}^i = (s_i \cdot \mathbf{u}^i) \oplus G(\mathbf{k}_i^{s_i})$. (Note that $\mathbf{q}^i = (s_i \cdot \mathbf{r}) \oplus \mathbf{t}^i$.)
- (c) Let $Q = [\mathbf{q}^1 | \dots | \mathbf{q}^\ell]$ denote the $m \times \ell$ bit matrix where its i th column is \mathbf{q}^i . Let \mathbf{q}_j denote the j th row of the matrix Q . (Note that $\mathbf{q}^i = (s_i \cdot \mathbf{r}) \oplus \mathbf{t}^i$ and $\mathbf{q}_j = (r_j \cdot \mathbf{s}) \oplus \mathbf{t}_j$.)
- (d) P_S sends (y_j^0, y_j^1) for every $1 \leq j \leq m$, where:

$$y_j^0 = x_j^0 \oplus H(j, \mathbf{q}_j) \quad \text{and} \quad y_j^1 = x_j^1 \oplus H(j, \mathbf{q}_j \oplus \mathbf{s})$$

- (e) For $1 \leq j \leq m$, P_R computes $x_j = y_j^{r_j} \oplus H(j, \mathbf{t}_j)$.

3. **Output:** P_R outputs $(x_1^{r_1}, \dots, x_m^{r_m})$; P_S has no output.

3 Our Protocol

In order to achieve security for malicious adversaries, we add a consistency check of the values \mathbf{r} that are sent in Step 2a of the semi-honest protocol (Protocol 1), and we increase the number of base-OTs. Let $\mathbf{r}^i = \mathbf{t}^i \oplus G(\mathbf{k}_i^1) \oplus \mathbf{u}^i$, i.e., the value that is implicitly defined by \mathbf{u}^i . We observe that if the receiver P_R uses the same choice bits \mathbf{r}^i and \mathbf{r}^j for some distinct $i, j \in [\ell]^2$, they cancel out when computing their XOR, i.e., $\mathbf{u}^i \oplus \mathbf{u}^j = (\mathbf{t}^i \oplus G(\mathbf{k}_i^1) \oplus \mathbf{r}^i) \oplus (\mathbf{t}^j \oplus G(\mathbf{k}_j^1) \oplus \mathbf{r}^j) = G(\mathbf{k}_i^0) \oplus G(\mathbf{k}_i^1) \oplus G(\mathbf{k}_j^0) \oplus G(\mathbf{k}_j^1)$. After the base-OTs, P_S holds $G(\mathbf{k}_i^{s_i})$ and $G(\mathbf{k}_j^{s_j})$ and in Step 2a of Protocol 1, P_R computes and sends $\mathbf{u}^i = G(\mathbf{k}_i^0) \oplus G(\mathbf{k}_i^1) \oplus \mathbf{r}^i$ and $\mathbf{u}^j = G(\mathbf{k}_j^0) \oplus G(\mathbf{k}_j^1) \oplus \mathbf{r}^j$. Now note that P_S can compute the XOR of the strings he received in the base-OTs $G(\mathbf{k}_i^{s_i}) \oplus G(\mathbf{k}_j^{s_j})$ as well as the “inverse” XOR of the strings received in the base-OTs $G(\mathbf{k}_i^{s_i}) \oplus G(\mathbf{k}_j^{s_j}) = G(\mathbf{k}_i^{s_i}) \oplus G(\mathbf{k}_j^{s_j}) \oplus \mathbf{u}^i \oplus \mathbf{u}^j$ if and only if P_R has correctly used $\mathbf{r}^i = \mathbf{r}^j$. However, P_S cannot check whether the “inverse” XOR

is correct, since it has no information about $G(\mathbf{k}_i^{\overline{s_i}})$ and $G(\mathbf{k}_j^{\overline{s_j}})$ (this is due to the security of the base-OTs that guarantees that P_S receives the keys $\mathbf{k}_i^{s_i}, \mathbf{k}_i^{\overline{s_i}}$ only, and learns nothing about $\mathbf{k}_i^{\overline{s_i}}, \mathbf{k}_j^{\overline{s_j}}$). P_R cannot give these values to P_S since this will totally break the secrecy of its choice bits. However, the P_R can send the hashes of these inverse values. Specifically, the P_R commits to the XORs of all strings $h_{i,j}^{p,q} = H(G(\mathbf{k}_i^p) \oplus G(\mathbf{k}_j^q))$, for all combinations of $p, q \in \{0, 1\}$. Now, given $h_{i,j}^{s_i, s_j}, h_{i,j}^{\overline{s_i}, \overline{s_j}}$, P_S checks that $h_{i,j}^{s_i, s_j} = H(G(\mathbf{k}_i^{s_i}) \oplus G(\mathbf{k}_j^{s_j}))$, and that $h_{i,j}^{\overline{s_i}, \overline{s_j}} = H(G(\mathbf{k}_i^{s_i}) \oplus G(\mathbf{k}_i^{s_j}) \oplus \mathbf{u}^i \oplus \mathbf{u}^j) = H(G(\mathbf{k}_i^{\overline{s_i}}) \oplus G(\mathbf{k}_j^{\overline{s_j}}))$. This check passes if $\mathbf{r}^i = \mathbf{r}^j$ and $h_{i,j}^{p,q}$ were set correctly.

If a malicious P_R tries to cheat and has chosen $\mathbf{r}^i \neq \mathbf{r}^j$, it has to convince P_S by computing $h_{i,j}^{p,q} = H(G(\mathbf{k}_i^p) \oplus G(\mathbf{k}_j^q) \oplus \mathbf{r}^i \oplus \mathbf{r}^j)$ for all $p, q \in \{0, 1\}$. However, P_S can check the validity of $h_{i,j}^{s_i, s_j} = H(G(\mathbf{k}_i^{s_i}) \oplus G(\mathbf{k}_j^{s_j}))$ while P_R remains oblivious to s_i, s_j . Hence, P_R can only convince P_S by guessing s_i, s_j , computing $h_{i,j}^{s_i, s_j}$ correctly and set $h_{i,j}^{\overline{s_i}, \overline{s_j}} = H(G(\mathbf{k}_i^{\overline{s_i}}) \oplus G(\mathbf{k}_j^{\overline{s_j}}) \oplus \mathbf{r}^i \oplus \mathbf{r}^j)$, which P_R cannot do better than with probability $1/2$. This means that P_R can only successfully learn ρ bits but will be caught except with probability $2^{-\rho}$. The full description of our new protocol is given in Protocol 2. We give some more explanations regarding the possibility of the adversary to cheat during the consistency check in §3.1.

We note that learning few bits of the secret \mathbf{s} does not directly break the security of the protocol once $|\mathbf{s}| > \kappa$. In particular, the values $\{H(\mathbf{t}_j \oplus \mathbf{s})\}_j$ are used to mask the inputs $\{x_j^{1-r_j}\}_j$. Therefore, when H is modelled as a random oracle and enough bits of \mathbf{s} remain hidden from the adversary, each value $H(\mathbf{t}_j \oplus \mathbf{s})$ is random, and the adversary cannot learn the input $x_j^{1-r_j}$. For simplicity we first prove security of our protocol in the random-oracle model. We later show that H can be replaced with a variant of a correlation-robustness assumption.

The advantage of our protocol over [NNOB12] is that P_S does not need to reveal any information about s_i, s_j when checking the consistency between r^i and r^j (as long as P_R does not cheat, in which case it risks getting caught). Hence, it can force P_R to check that \mathbf{r}^i equals any \mathbf{r}^j , for $1 \leq j \leq \ell$ without disclosing any information.

Section outline. In the following, we describe our basic protocol and prove its security (§3.1). We then show how to reduce the number of consistency checks to achieve better performance (§3.2), and how to replace the random oracle with a weaker correlation robustness assumption (§3.3). Finally, we show how our protocol can be used to achieve covert security (§3.4).

PROTOCOL 2 (Our active-secure OT extension protocol)

- **Input of P_S :** m pairs (x_j^0, x_j^1) of n -bit strings, $1 \leq j \leq m$.
- **Input of P_R :** m selection bits $\mathbf{r} = (r_1, \dots, r_m)$.
- **Common Input:** Symmetric security parameter κ and statistical security parameter ρ . It is assumed that $\ell = \kappa + \rho$.
- **Oracles and cryptographic primitives:** The parties use an ideal $\ell \times OT_\kappa$ functionality, pseudorandom generator $G : \{0, 1\}^\kappa \rightarrow \{0, 1\}^{m+\kappa}$, and random-oracle H (see §3.3 for instantiation of H .)

1. *Initial OT Phase:*

- (a) P_S initializes a random vector $\mathbf{s} = (s_1, \dots, s_\ell) \in \{0, 1\}^\ell$ and P_R chooses ℓ pairs of seeds $\mathbf{k}_i^0, \mathbf{k}_i^1$ each of size κ .
- (b) The parties invoke the $\ell \times OT_\kappa$ -functionality, where P_S acts as the *receiver* with input \mathbf{s} and P_R acts as the *sender* with inputs $(\mathbf{k}_i^0, \mathbf{k}_i^1)$ for every $1 \leq i \leq \ell$.

For every $1 \leq i \leq \ell$, let $\mathbf{t}^i = G(\mathbf{k}_i^0)$. Let $T = [\mathbf{t}^1 | \dots | \mathbf{t}^\ell]$ denote the $(m + \kappa) \times \ell$ bit matrix where its i th column is \mathbf{t}^i for $1 \leq i \leq \ell$. Let \mathbf{t}_j denote the j th row of T for $1 \leq j \leq m$.

2. *OT Extension Phase (Part I):*

- (a) P_R chooses a random string $\boldsymbol{\tau} \in \{0, 1\}^\kappa$, and defines $\mathbf{r}' = \mathbf{r} || \boldsymbol{\tau}$.
- (b) P_R computes $\mathbf{t}^i = G(\mathbf{k}_i^0)$ and $\mathbf{u}^i = \mathbf{t}^i \oplus G(\mathbf{k}_i^1) \oplus \mathbf{r}'$, and sends \mathbf{u}^i to P_S for every $1 \leq i \leq \ell$.

3. *Consistency Check of \mathbf{r}' :* (the main change from Protocol 1)

- (a) For every pair $\alpha, \beta \subseteq [\ell]^2$, P_R defines the four values:

$$\begin{aligned} h_{\alpha, \beta}^{0,0} &= H(G(\mathbf{k}_\alpha^0) \oplus G(\mathbf{k}_\beta^0)) & h_{\alpha, \beta}^{0,1} &= H(G(\mathbf{k}_\alpha^0) \oplus G(\mathbf{k}_\beta^1)), \\ h_{\alpha, \beta}^{1,0} &= H(G(\mathbf{k}_\alpha^1) \oplus G(\mathbf{k}_\beta^0)) & h_{\alpha, \beta}^{1,1} &= H(G(\mathbf{k}_\alpha^1) \oplus G(\mathbf{k}_\beta^1)). \end{aligned}$$

It then sends $\mathcal{H}_{\alpha, \beta} = (h_{\alpha, \beta}^{0,0}, h_{\alpha, \beta}^{0,1}, h_{\alpha, \beta}^{1,0}, h_{\alpha, \beta}^{1,1})$ to P_S .

- (b) For every pair $\alpha, \beta \subseteq [\ell]^2$, P_S knows $s_\alpha, s_\beta, \mathbf{k}_\alpha^{s_\alpha}, \mathbf{k}_\beta^{s_\beta}, \mathbf{u}^\alpha, \mathbf{u}^\beta$ and checks that:

- i. $h_{\alpha, \beta}^{s_\alpha, s_\beta} = H(G(\mathbf{k}_\alpha^{s_\alpha}) \oplus G(\mathbf{k}_\beta^{s_\beta}))$.
- ii. $h_{\alpha, \beta}^{\overline{s_\alpha}, \overline{s_\beta}} = H(G(\mathbf{k}_\alpha^{s_\alpha}) \oplus G(\mathbf{k}_\beta^{s_\beta}) \oplus \mathbf{u}^\alpha \oplus \mathbf{u}^\beta) \quad (= H(G(\mathbf{k}_\alpha^{\overline{s_\alpha}}) \oplus G(\mathbf{k}_\beta^{\overline{s_\beta}}) \oplus \mathbf{r}^\alpha \oplus \mathbf{r}^\beta))$.
- iii. $\mathbf{u}^\alpha \neq \mathbf{u}^\beta$.

In case one of these checks fails, P_S aborts and outputs \perp .

4. *OT Extension Phase (Part II):*

- (a) For every $1 \leq i \leq \ell$, P_S defines $\mathbf{q}^i = (s_i \cdot \mathbf{u}^i) \oplus G(\mathbf{k}_i^{s_i})$. (Note that $\mathbf{q}^i = (s_i \cdot \mathbf{r}) \oplus \mathbf{t}^i$.)
- (b) Let $Q = [\mathbf{q}^1 | \dots | \mathbf{q}^\ell]$ denote the $(m + \kappa) \times \ell$ bit matrix where its i th column is \mathbf{q}^i . Let \mathbf{q}_j denote the j th row of the matrix Q . (Note that $\mathbf{q}^i = (s_i \cdot \mathbf{r}) \oplus \mathbf{t}^i$ and $\mathbf{q}_j = (r_j \cdot \mathbf{s}) \oplus \mathbf{t}_j$.)
- (c) P_S sends (y_j^0, y_j^1) for every $1 \leq j \leq m$, where:

$$y_j^0 = x_j^0 \oplus H(j, \mathbf{q}_j) \quad \text{and} \quad y_j^1 = x_j^1 \oplus H(j, \mathbf{q}_j \oplus \mathbf{s})$$

- (d) For $1 \leq j \leq m$, P_R computes $x_j = y_j^{r_j} \oplus H(j, \mathbf{t}_j)$.

5. **Output:** P_R outputs $(x_1^{r_1}, \dots, x_m^{r_m})$; P_S has no output.

3.1 The Security of Our Protocol

Malicious sender. The original OT extension protocol of [IKNP03] already provides security against a malicious P_S . In an earlier version of this paper, we claimed that our checks in Step 3 of the protocol do not provide a malicious P_S with any new capabilities, and we derived security directly from the semi-honest case with no further proof.

However, as pointed out by [Sch], adding the checks for consistency of \mathbf{r} provides an “oracle” for checking whether a particular candidate $\tilde{\mathbf{r}}$ is the input of the receiver or not. In fact, given \mathbf{u}^α , $G(\mathbf{k}_\alpha^{s_\alpha})$ and $G(\mathbf{k}_\beta^{s_\beta})$ (note that $\mathbf{k}_\alpha^{s_\alpha}$ and $\mathbf{k}_\beta^{s_\beta}$ are sent to the sender during the base OT stage), the malicious sender can check whether $\mathbf{r} = \tilde{\mathbf{r}}$ or not, by first computing $\overline{\mathbf{t}}_\alpha^{s_\alpha} = \mathbf{u}^\alpha \oplus G(\mathbf{k}_\alpha^{s_\alpha}) \oplus \tilde{\mathbf{r}}$ and then checking whether $h_{\alpha,\beta}^{\overline{s_\alpha}, s_\beta} = H(\overline{\mathbf{t}}_\alpha^{s_\alpha} \oplus G(\mathbf{k}_\beta^{s_\beta}))$.

In order to prevent this subtle attack, we modify the protocol such that P_R first appends to \mathbf{r} some random string $\boldsymbol{\tau} \in \{0, 1\}^\kappa$ resulting in a string $\mathbf{r}' = \mathbf{r} \parallel \boldsymbol{\tau}$ (this is Step 2a in the protocol). This adds entropy to the choice of \mathbf{r} and therefore prevents the above attack. Note that in the second part of the OT extension phase, the transfer is done on only the first m rows of the matrix Q and not $m + \kappa$. In Appendix C we formally prove that the protocol is secure in the presence of a malicious sender, assuming that the function H is modeled as κ -min-entropy strongly correlation robust (see §3.10 regarding this assumption). We also note that the change in the efficiency of the protocol is minor, as $m \gg \kappa$.

Simulating a malicious receiver. In the case of a malicious receiver, the adversary may not use the same \mathbf{r} in the messages $\mathbf{u}^1, \dots, \mathbf{u}^\ell$, and as a result learn some bits from the secret \mathbf{s} . Therefore, we add a consistency check of \mathbf{r} to the semi-honest protocol of [IKNP03]. However, this verification of consistency of \mathbf{r} is not perfectly sound, and the verification may still pass even when the receiver sends few \mathbf{u} 's that do not define the same \mathbf{r} . This makes the analysis a bit more complicated.

For every $1 \leq i \leq \ell$, let $\mathbf{r}^i \stackrel{\text{def}}{=} \mathbf{u}^i \oplus G(\mathbf{k}_i^0) \oplus G(\mathbf{k}_i^1)$ that is, the “input” \mathbf{r}^i which is implicitly defined by \mathbf{u}^i and the base-OTs.

We now explore how the matrices Q, T are changed when the adversary uses inconsistent \mathbf{r} 's. Recall that when the receiver uses the same \mathbf{r} , then $\mathbf{q}^i = (s_i \cdot \mathbf{r}) \oplus \mathbf{t}^i$ and $\mathbf{q}_j = (r_j \cdot \mathbf{s}) \oplus \mathbf{t}_j$. However, in case of inconsistent \mathbf{r} 's, we get that $\mathbf{q}^i = (s_i \cdot \mathbf{r}^i) \oplus \mathbf{t}^i$. The case of \mathbf{q}_j is rather more involved; let $R = [\mathbf{r}^1 \mid \dots \mid \mathbf{r}^\ell]$ denote the $m \times \ell$ matrix where its i th column is \mathbf{r}^i , and let \mathbf{r}_j denote the j th row of the matrix R . For two strings of the same length $\mathbf{a} = (a_1, \dots, a_k), \mathbf{b} = (b_1, \dots, b_k)$, let $\mathbf{a} * \mathbf{b}$ define the entry-wise product, that is $\mathbf{a} * \mathbf{b} = (a_1 \cdot b_1, \dots, a_k \cdot b_k)$. We get that $\mathbf{q}_j = (\mathbf{r}_j * \mathbf{s}) \oplus \mathbf{t}_j$ (note that in an honest execution, \mathbf{r}_j is the same bit everywhere). The sender masks the inputs (x_j^0, x_j^1) with $(H(j, \mathbf{q}_j), H(j, \mathbf{q}_j \oplus \mathbf{s}))$.

In order to understand better the value \mathbf{q}_j , let $\mathbf{r} = (r_1, \dots, r_m)$ be the string that occurs the most from the set $\{\mathbf{r}^1, \dots, \mathbf{r}^\ell\}$, and let $\mathcal{U} \subset [\ell]$ be the set of all indices for which $\mathbf{r}^i = \mathbf{r}$ for all $i \in \mathcal{U}$. Let $B = [\ell] \setminus \mathcal{U}$ be the complementary set, that is, the set of all indices for which for every $i \in B$ it holds that $\mathbf{r}^i \neq \mathbf{r}$. As we will see below, except with some negligible probability, the verification phase guarantees that $|\mathcal{U}| \geq \ell - \rho$. Thus, for every $1 \leq j \leq m$, the vector \mathbf{r}_j (which is the j th row of the matrix R), can be represented as $\mathbf{r}_j = (r_j \cdot \mathbf{1}) \oplus \mathbf{e}_j$, where $\mathbf{1}$ is the all one vector of size ℓ , and \mathbf{e}_j is some error vector with Hamming distance at most ρ from $\mathbf{0}$. Note that the non-zero indices in \mathbf{e}_j are all in B . Thus, we conclude that:

$$\mathbf{q}_j = (\mathbf{s} * \mathbf{r}_j) \oplus \mathbf{t}_j = (\mathbf{s} * (r_j \cdot \mathbf{1} \oplus \mathbf{e}_j)) \oplus \mathbf{t}_j = (r_j \cdot \mathbf{s}) \oplus \mathbf{t}_j \oplus (\mathbf{s} * \mathbf{e}_j).$$

Recall that in an honest execution $\mathbf{q}_j = (r_j \cdot \mathbf{s}) \oplus \mathbf{t}_j$, and therefore the only difference is the term $(\mathbf{s} * \mathbf{e}_j)$. Moreover, note that $\mathbf{s} * \mathbf{e}_j$ completely hides all the bits of \mathbf{s} that are in \mathcal{U} , and may expose only the bits that are in B . Thus, the consistency check of \mathbf{r} guarantees two important properties: First, that almost all the inputs are consistent with some implicitly defined string \mathbf{r} , and thus the bits r_j are uniquely defined. Second, the set of inconsistent inputs (i.e., the set B) is small, and thus the adversary may learn only a limited amount of bits of \mathbf{s} .

The consistency checks of \mathbf{r} . We now examine what properties are guaranteed by our consistency check, for a single pair (α, β) . The malicious receiver P_R first sends the set of keys $\mathcal{K} = \{\mathbf{k}_i^0, \mathbf{k}_i^1\}$ to the base-OT protocol, and then sends all the values $(\mathbf{u}^1, \dots, \mathbf{u}^\ell)$ and the checks $\mathcal{H} = \{\mathcal{H}_{\alpha, \beta}\}_{\alpha, \beta}$. In the simulation, the simulator can choose \mathbf{s} only after it receives all these messages (this is because the adversary gets no output from the invocation of the OT primitive). Thus, for a given set of messages that the adversary outputs, we can ask what is the number of secrets \mathbf{s} for which the verification will pass, and the number for which it will fail. If the verification passes for some given $\mathcal{T} = (\mathcal{K}, \mathbf{u}^1, \dots, \mathbf{u}^\ell, \mathcal{H})$ and some secret \mathbf{s} , then we say that \mathcal{T} is consistent with \mathbf{s} ; In case the verification fails, we say that \mathcal{T} is inconsistent.

The following Lemma considers the values that the adversary has sent regarding some pair (α, β) , and considers the relation to the pair of bits (s_α, s_β) of the secret \mathbf{s} . We have:

Lemma 3.1 *Let $\mathcal{T}_{\alpha, \beta} = \{\mathcal{H}_{\alpha, \beta}, \mathbf{u}^\alpha, \mathbf{u}^\beta, \{\mathbf{k}_\alpha^b\}_b, \{\mathbf{k}_\beta^b\}_b\}$ and assume that H is a collision-resistant hash-function. We have:*

1. *If $\mathbf{r}^\alpha \neq \mathbf{r}^\beta$ and $\mathcal{T}_{\alpha, \beta}$ is consistent with (s_α, s_β) , then it is inconsistent with $(\overline{s_\alpha}, \overline{s_\beta})$.*
2. *If $\mathbf{r}^\alpha = \mathbf{r}^\beta$ and $\mathcal{T}_{\alpha, \beta}$ is consistent with (s_α, s_β) , then it is consistent also with $(\overline{s_\alpha}, \overline{s_\beta})$.*

Proof: For the first item, assume that $\mathbf{r}^\alpha \neq \mathbf{r}^\beta$ and that $\mathcal{T}_{\alpha, \beta}$ is consistent both with (s_α, s_β) and $(\overline{s_\alpha}, \overline{s_\beta})$. Thus, from the check of consistency of (s_α, s_β) :

$$h_{\alpha, \beta}^{s_\alpha, s_\beta} = H\left(G(\mathbf{k}_\alpha^{s_\alpha}) \oplus G(\mathbf{k}_\beta^{s_\beta})\right), \quad h_{\alpha, \beta}^{\overline{s_\alpha}, \overline{s_\beta}} = H\left(G(\mathbf{k}_\alpha^{s_\alpha}) \oplus G(\mathbf{k}_\beta^{s_\beta}) \oplus \mathbf{u}^\alpha \oplus \mathbf{u}^\beta\right),$$

and that $\mathbf{u}^\alpha \neq \mathbf{u}^\beta$. In addition, from the check of consistency of $(\overline{s_\alpha}, \overline{s_\beta})$ it holds that:

$$h_{\alpha, \beta}^{\overline{s_\alpha}, \overline{s_\beta}} = H\left(G(\mathbf{k}_\alpha^{\overline{s_\alpha}}) \oplus G(\mathbf{k}_\beta^{\overline{s_\beta}})\right), \quad h_{\alpha, \beta}^{s_\alpha, s_\beta} = H\left(G(\mathbf{k}_\alpha^{\overline{s_\alpha}}) \oplus G(\mathbf{k}_\beta^{\overline{s_\beta}}) \oplus \mathbf{u}^\alpha \oplus \mathbf{u}^\beta\right),$$

and that $\mathbf{u}^\alpha \neq \mathbf{u}^\beta$. This implies that:

$$H\left(G(\mathbf{k}_\alpha^{s_\alpha}) \oplus G(\mathbf{k}_\beta^{s_\beta})\right) = h_{\alpha, \beta}^{s_\alpha, s_\beta} = H\left(G(\mathbf{k}_\alpha^{\overline{s_\alpha}}) \oplus G(\mathbf{k}_\beta^{\overline{s_\beta}}) \oplus \mathbf{u}^\alpha \oplus \mathbf{u}^\beta\right),$$

and from the collision resistance property of H we get that:

$$G(\mathbf{k}_\alpha^{s_\alpha}) \oplus G(\mathbf{k}_\beta^{s_\beta}) = G(\mathbf{k}_\alpha^{\overline{s_\alpha}}) \oplus G(\mathbf{k}_\beta^{\overline{s_\beta}}) \oplus \mathbf{u}^\alpha \oplus \mathbf{u}^\beta.$$

Recall that $\mathbf{r}^\alpha \stackrel{\text{def}}{=} \mathbf{u}^\alpha \oplus G(\mathbf{k}_\alpha^0) \oplus G(\mathbf{k}_\alpha^1)$, and $\mathbf{r}^\beta \stackrel{\text{def}}{=} \mathbf{u}^\beta \oplus G(\mathbf{k}_\beta^0) \oplus G(\mathbf{k}_\beta^1)$. Combining the above, we get that $\mathbf{r}^\alpha = \mathbf{r}^\beta$, in contradiction.

For the second item, once $\mathbf{r}^\alpha = \mathbf{r}^\beta$, we get that $\mathbf{u}^\alpha \oplus \mathbf{u}^\beta = G(\mathbf{k}_\alpha^0) \oplus G(\mathbf{k}_\alpha^1) \oplus G(\mathbf{k}_\beta^0) \oplus G(\mathbf{k}_\beta^1)$ and it is easy to see that if the consistency check of (s_α, s_β) holds, then the consistency check of $(\overline{s_\alpha}, \overline{s_\beta})$ holds also. ■

Lemma 3.1 implies what attacks the adversary can do, and what bits of \mathbf{s} it can learn from each such an attack. In the following, we consider a given partial transcript $\mathcal{T}_{\alpha,\beta} = ((\mathbf{k}_\alpha^0, \mathbf{k}_\alpha^1, \mathbf{k}_\beta^0, \mathbf{k}_\beta^1), (\mathbf{u}^\alpha, \mathbf{u}^\beta), \mathcal{H}_{\alpha,\beta})$ and analyze what the messages might be, and what the adversary learns in case the verification passes. Let $\mathbf{r}^\alpha = \mathbf{u}^\alpha \oplus G(\mathbf{k}_\alpha^0) \oplus G(\mathbf{k}_\alpha^1)$ and \mathbf{r}^β defined analogously. We consider 4 types:

1. **Type 1: correct.** In this case, it holds that $\mathbf{r}^\alpha = \mathbf{r}^\beta$, and for every $(a, b) \in \{0, 1\}^2$: $h_{\alpha,\beta}^{a,b} = H(G(\mathbf{k}_\alpha^a) \oplus G(\mathbf{k}_\beta^b))$. The verification passes for every possible value of (s_α, s_β) .

2. **Type 2: $\mathbf{r}^\alpha = \mathbf{r}^\beta$, but $\mathcal{H}_{\alpha,\beta}$ is incorrect.** In this case, the adversary sent $\mathbf{u}^\alpha, \mathbf{u}^\beta$ that define the same \mathbf{r} . However, it may send hashes $\mathcal{H}_{\alpha,\beta}$ that are incorrect (i.e., for some $(a, b) \in \{0, 1\}^2$, it may send: $h_{\alpha,\beta}^{a,b} \neq H(G(\mathbf{k}_\alpha^a) \oplus G(\mathbf{k}_\beta^b))$). However, from Lemma 3.1, if $\mathbf{r}^\alpha = \mathbf{r}^\beta$ and $\mathcal{H}_{\alpha,\beta}$ is consistent with (s_α, s_β) then it is also consistent with $(\overline{s_\alpha}, \overline{s_\beta})$.

Thus, a possible attack of the adversary, for instance, is to send correct hashes for some bits (0,0) and (1,1), but incorrect ones for (0,1) and (1,0). The verification will pass with probability 1/2, exactly if (s_α, s_β) are either (0,0) or (1,1), but it will fail in the other two cases (i.e., (1,0) or (0,1)). We therefore conclude that the adversary may learn the relation $s_\alpha \oplus s_\beta$, and gets caught with probability 1/2.

3. **Type 3: $\mathbf{r}^\alpha \neq \mathbf{r}^\beta$ and $\mathcal{H}_{\alpha,\beta}$ is incorrect in two positions.** In this case, for instance, the adversary can set the values $h_{\alpha,\beta}^{0,0}, h_{\alpha,\beta}^{0,1}$ correctly (i.e., $h_{\alpha,\beta}^{0,0} = H(G(\mathbf{k}_\alpha^0) \oplus G(\mathbf{k}_\beta^0))$ and $h_{\alpha,\beta}^{0,1} = H(G(\mathbf{k}_\alpha^0) \oplus G(\mathbf{k}_\beta^1))$) and set values $h_{\alpha,\beta}^{1,0}, h_{\alpha,\beta}^{1,1}$ accordingly, such that the verification will pass for the cases of $(s_\alpha, s_\beta) = (0, 0)$ or $(0, 1)$. That is, it sets:

$$h_{\alpha,\beta}^{1,0} = H(G(\mathbf{k}_\alpha^0) \oplus G(\mathbf{k}_\beta^1) \oplus \mathbf{u}^\alpha \oplus \mathbf{u}^\beta)$$

(and it sets $h_{\alpha,\beta}^{1,1}$ in a similar way). In this case, the adversary succeeds with probability 1/2 and learns that $s_\alpha = 0$ in case the verification passes. Similarly, it can guess the value of s_β and set the values accordingly. For conclusion, the adversary can learn whether its guess was correct, and in which case it learns exactly one of the bits s_α or s_β but does not learn anything about the other bit.

4. **Type 4: $\mathbf{r}^\alpha \neq \mathbf{r}^\beta$ and $\mathcal{H}_{\alpha,\beta}$ is incorrect in three positions.** In this case, the adversary may guess both bits $(s_\alpha, s_\beta) = (a, b)$ and set $h_{\alpha,\beta}^{a,b}$ correctly, set $h_{\alpha,\beta}^{\overline{a},\overline{b}}$ accordingly (i.e., such that the verification will pass for (a, b)), but will fail for any one of the other cases. In this case, the adversary learns the values (s_α, s_β) entirely, but succeeds with probability of at most 1/4.

Note that whenever $\mathbf{r}^\alpha \neq \mathbf{r}^\beta$, the adversary may pass the verification of the pair (α, β) with probability of at most 1/2. This is because it cannot send consistent hashes for all possible values of (s_α, s_β) , and must, in some sense, “guess” either one of the bits, or both (i.e., Type 3 or Type 4). However, an important point that makes the analysis more difficult is the fact that the two checks are not necessarily independent. That is, in case where $\mathbf{r}^\alpha \neq \mathbf{r}^\beta$ and $\mathbf{r}^\beta \neq \mathbf{r}^\gamma$, although the probability to pass each one of the verification of (α, β) and (β, γ) separately is at most 1/2, the probability to pass both verifications together is higher than 1/4, and these two checks are not independent. This is because the adversary can guess the bit s_β , and set the hashes as in Type 3

in both checks. The adversary will pass these two checks if it guesses s_β correctly, with probability $1/2$.

Proving security of the protocol. Before proceeding to the full proof of security, we first sketch the proof. The simulator \mathcal{S} invokes the malicious receiver and plays the role of the base-OT trusted party and the honest sender. It receives from the adversary its inputs to the base-OTs, and thus knows the values $\{\mathbf{k}_i^0, \mathbf{k}_i^1\}_{i=1}^\ell$. Therefore, it can compute all the values $\mathbf{r}^1, \dots, \mathbf{r}^\ell$ when it receives the messages $\mathbf{u}^1, \dots, \mathbf{u}^\ell$. It computes the set of indices \mathcal{U} , and extracts \mathbf{r} . It then performs the same checks as an honest sender, in Step 3 of Protocol 2, and aborts the execution if the adversary is caught cheating. Then, it sends the trusted party the value \mathbf{r} that it has extracted, and learns the inputs $x_1^{r_1}, \dots, x_m^{r_m}$. It computes \mathbf{q}_j as instructed in the protocol (recall that these \mathbf{q}_j may contain the additional “shift” $\mathbf{s} * \mathbf{e}_j$) and use some random values for all $\{y_j^{\overline{r_j}}\}_{j=1}^m$.

Since the values $\{y_j^{\overline{r_j}}\}_{j=1}^m$ are random in the ideal execution, and equal $\{x_j^{\overline{r_j}} \oplus H(j, \mathbf{q}_j \oplus \mathbf{s})\}$ in the real execution, a distinguisher may distinguish between the real and ideal execution once it makes a query of the form $(j, \mathbf{q}_j \oplus \mathbf{s})$ to the random oracle. We claim, however, that the probability that the distinguisher will make such a query is bounded by $(t+1)/|S|$, where t is the number of queries it makes to the random oracle, and S is the set of all possible secrets \mathbf{s} that are consistent with the view that it receives. Thus, once we show that $|S| > 2^\kappa$, the probability that it will distinguish between the real and ideal execution is negligible in κ .

However, the above description is too simplified. First, if the adversary performs few attacks of Type 2, it learns information regarding \mathbf{s} from the mere fact that the verification has passed. Moreover, recall that $y_j^{\overline{r_j}} = x_j^{\overline{r_j}} \oplus H(j, \mathbf{t}_j \oplus (\mathbf{s} * \mathbf{e}_j))$, and that the adversary can control the values \mathbf{t}_j and \mathbf{e}_j . Recall that \mathbf{e}_j is a vector that is all zero in positions that are in \mathcal{U} , and may vary in positions that are in B . This implies that by simple queries to the random oracle, and by choosing the vectors \mathbf{e}_j cleverly, the adversary can totally reveal the bits \mathbf{s}_B quite easily. We therefore have to show that the set B is small, while also showing that the set of consistent secrets is greater than 2^κ (that is, $|S| \geq 2^\kappa$). We now proceed to a formal statement of the theorem and formal proof of security, where there we prove the two informal claims that were just mentioned.

Theorem 3.2 *Assuming that H is a random oracle, G is a pseudo-random generator, Protocol 2 with $\ell = \kappa + \rho$ securely computes the $m \times OT_n$ functionality in the $\ell \times OT_\kappa$ -hybrid model in the presence of a static malicious adversary, where κ is the symmetric security parameter and ρ is the statistical security parameter.*

Proof: Since we already gave some proof sketch, we start directly with a formal description of the simulator \mathcal{S} :

The simulator \mathcal{S} .

1. The simulator invokes the adversary \mathcal{A} on the auxiliary input z .
2. **Initial OT phase:** The adversary \mathcal{A} outputs ℓ pairs of κ -bits each $\{\mathbf{k}_i^0, \mathbf{k}_i^1\}_{i=1}^\ell$ as input to the $\ell \times OT_\kappa$ -functionality. It receives no output from this invocation.
3. **First part of OT extension phase:** The adversary \mathcal{A} outputs ℓ strings $\mathbf{u}^1, \dots, \mathbf{u}^\ell$.
4. **Consistency check of \mathbf{r} :**
 - (a) For every $\alpha, \beta \in [\ell]^2$, the adversary \mathcal{A} outputs the quadruple $H^{\alpha, \beta} = (h_{\alpha, \beta}^{0,0}, h_{\alpha, \beta}^{0,1}, h_{\alpha, \beta}^{1,0}, h_{\alpha, \beta}^{1,1})$.

- (b) The simulator chooses a string \mathbf{s} uniformly at random from $\{0, 1\}^\ell$.
- (c) Given the values $\{\{\mathbf{k}_i^0, \mathbf{k}_i^1\}_{i=1}^\ell, \mathbf{u}^1, \dots, \mathbf{u}^\ell, \{H^{\alpha, \beta}\}_{\alpha, \beta}\}$ and the chosen secret \mathbf{s} , the simulator can perform all the needed checks as the honest sender in the real execution. In case where one of the verification fails, the simulator halts.

5. Second part of the OT extension phase:

- (a) The simulator computes the matrices T , Q and R , where for every i , $\mathbf{t}^i = G(\mathbf{k}_i^0)$, $\mathbf{q}^i = (s_i \cdot \mathbf{r}^i) \oplus \mathbf{t}^i$ and $\mathbf{r}^i = G(\mathbf{k}_i^0) \oplus G(\mathbf{k}_i^1) \oplus \mathbf{u}^i$.
- (b) From all the vectors $\mathbf{r}^1, \dots, \mathbf{r}^\ell$, let \mathbf{r} be the vector that is mostly repeated (as we will see, the verification process guarantees that there exists a vector that is repeated at least $\ell - \rho$ times).
- (c) Send \mathbf{r} to the trusted party, and receive $x_1^{r_1}, \dots, x_m^{r_m}$. Define the values \mathbf{e}_j for every $1 \leq j \leq m$ (explicitly, define the matrix R as the matrix for which its i th column is \mathbf{r}^i , and let \mathbf{r}_j denote its j th row. Then, $\mathbf{e}_j = (r_j \cdot \mathbf{1}) \oplus \mathbf{r}_j$. Then, for every $1 \leq j \leq m$, set $y_j^{r_j} = x_j^{r_j} \oplus H(j, \mathbf{t}_j \oplus (\mathbf{s} * \mathbf{e}_j))$, and set $y_j^{r_j}$ uniformly at random. Send $\{(y_j^0, y_j^1)\}_{j=1}^m$ to the adversary \mathcal{A} , output whatever it outputs and halt.

Let $\mathcal{T} = \{\{\mathbf{k}_i^0, \mathbf{k}_i^1\}_{i=1}^\ell, \mathbf{u}^1, \dots, \mathbf{u}^\ell, \{H_{\alpha, \beta}\}_{\alpha, \beta}\}$, i.e., the values that the adversary gives during the execution of the protocol. Observe that the simulator chooses the secret \mathbf{s} only after \mathcal{T} is determined (since the adversary receives no output from the execution of the OT primitive, we can assume that). We divide all possible \mathcal{T} into two sets, $\mathcal{T}_{\text{good}}$ and \mathcal{T}_{bad} , defined as follows:

$$\mathcal{T}_{\text{good}} = \left\{ \mathcal{T} \mid \Pr_{\mathbf{s}} [\text{consistent}(\mathcal{T}, \mathbf{s}) = 1] > 2^{-\rho} \right\} \quad \text{and} \quad \mathcal{T}_{\text{bad}} = \left\{ \mathcal{T} \mid \Pr_{\mathbf{s}} [\text{consistent}(\mathcal{T}, \mathbf{s}) = 1] \leq 2^{-\rho} \right\}.$$

where $\text{consistent}(\mathcal{T}, \mathbf{s})$ is a predicate that gets 1 when the verification passes for the transcript \mathcal{T} and the secret \mathbf{s} , and 0 otherwise. The probability is taken over the choice of \mathbf{s} . For a given \mathcal{T} , let $\mathcal{S}(\mathcal{T})$ be the set of all possible secrets $\mathbf{s} \in \{0, 1\}^\ell$, that are consistent with \mathcal{T} . That is: $\mathcal{S}(\mathcal{T}) = \{\mathbf{s} \in \{0, 1\}^\ell \mid \text{consistent}(\mathcal{T}, \mathbf{s}) = 1\}$. Therefore, it holds that:

$$\Pr_{\mathbf{s}} [\text{consistent}(\mathcal{T}, \mathbf{s}) = 1] = \frac{|\mathcal{S}(\mathcal{T})|}{2^\ell}$$

and thus $|\mathcal{S}(\mathcal{T})| = 2^\ell \cdot \Pr [\text{consistent}(\mathcal{T}, \mathbf{s}) = 1]$. As a result, for every $\mathcal{T} \in \mathcal{T}_{\text{good}}$, it holds that $|\mathcal{S}(\mathcal{T})| > 2^\ell \cdot 2^{-\rho} = 2^{\ell - \rho} = 2^\kappa$. That is, in case a transcript $\mathcal{T} \in \mathcal{T}_{\text{good}}$ passes the consistency check of \mathbf{r} , there are at least 2^κ different secrets \mathbf{s} that are consistent with the given transcript, each are likely with the same probability, and thus the adversary may guess \mathbf{s} with probability of at most $2^{-\kappa}$.

Let \mathcal{U} be the largest set of indices such that for every $i, j \in \mathcal{U}$, $\mathbf{r}^i = \mathbf{r}^j$. Let B be the complementary set, that is, $B = [\ell] \setminus \mathcal{U}$. From the definition of the sets, for every $\alpha \in \mathcal{U}$ and $\beta \in B$, it holds that $\mathbf{r}^\alpha \neq \mathbf{r}^\beta$.

We claim that if $|\mathcal{U}| < \ell - \rho$ (i.e., $|B| > \rho$), then it must hold that $\mathcal{T} \in \mathcal{T}_{\text{bad}}$ and the adversary gets caught with high probability. That is:

Claim 3.3 *Let \mathcal{T} be as above, and let \mathcal{U} be the largest set of indices such that for every $\alpha, \beta \in \mathcal{U}$, $\mathbf{r}^\alpha = \mathbf{r}^\beta$. Assume that $|\mathcal{U}| < \ell - \rho$. Then:*

$$\Pr_{\mathbf{s}} [\text{consistent}(\mathcal{T}, \mathbf{s}) = 1] \leq 2^{-\rho}$$

and thus, $\mathcal{T} \in \mathcal{T}_{\text{bad}}$.

We will prove the claim below. Let $\mathcal{T} \in \mathcal{T}_{\text{good}}$, and let \mathcal{U} and B be as above. Using the claim above, we have that $|B| < \rho$. We now focus on the set of secrets \mathbf{s} that are consistent with this transcript \mathcal{T} , i.e., the set $S(\mathcal{T})$. We claim that the bits \mathbf{s}_B are common to all the secrets in $S(\mathcal{T})$, and thus, even when we give the adversary the bits \mathbf{s}_B in the clear once the verification is completed, the adversary still has to guess \mathbf{s} from a set of at least 2^κ secrets. Formally:

Claim 3.4 *Let $\mathcal{T} \in \mathcal{T}_{\text{good}}$, and let \mathcal{U} and B as above. Then, there exists a string $w \in \{0,1\}^{|B|}$, such that for every $\mathbf{s}' \in S(\mathcal{T})$, it holds that: $\mathbf{s}'_B = w$.*

Proof: From the definition of the sets B and \mathcal{U} , it holds that for every $\alpha \in \mathcal{U}$ and $\beta \in B$, $\mathbf{r}^\alpha \neq \mathbf{r}^\beta$. Consider two secrets \mathbf{s}, \mathbf{s}' that are consistent with \mathcal{T} (since $\mathcal{T} \in \mathcal{T}_{\text{good}}$, there are many such \mathcal{T} 's). We show the following:

- If $\mathbf{s}_B \neq \mathbf{s}'_B$ then $\mathbf{s}_\mathcal{U} = \mathbf{s}'_\mathcal{U}$.
- If $\mathbf{s}_\mathcal{U} \neq \mathbf{s}'_\mathcal{U}$ then $\mathbf{s}_B = \mathbf{s}'_B$.

We show the first claim. Assume that $\mathbf{s}_B \neq \mathbf{s}'_B$, thus, there must exist an index $\beta \in B$ such that $s_\beta \neq s'_\beta$, i.e., $s'_\beta = \bar{s}_\beta$. Now, consider some $\alpha \in \mathcal{U}$, we show that $s_\alpha = s'_\alpha$ and thus $\mathbf{s}_\mathcal{U} = \mathbf{s}'_\mathcal{U}$. From Lemma 3.1, since \mathcal{T} is consistent with (s_α, s_β) , it is inconsistent with $(\bar{s}_\alpha, \bar{s}_\beta) = (\bar{s}_\alpha, s'_\beta)$. Therefore, in order to have consistency between \mathbf{s}' and \mathcal{T} , we must have that $s'_\alpha \neq \bar{s}_\alpha$, and so $s'_\alpha = s_\alpha$. The second claim is proven analogously.

We now claim that the set $S(\mathcal{T})$ either shares the same \mathbf{s}_B for all its elements, or shares the same $\mathbf{s}_\mathcal{U}$ for all elements (and of course, not both). In order to see this, let $S(\mathcal{T}) = \{\mathbf{s}^1, \dots, \mathbf{s}^n\}$. Assume, without loss of generality, that $\mathbf{s}^1_\mathcal{U} \neq \mathbf{s}^2_\mathcal{U}$ (and so, $\mathbf{s}^1_B = \mathbf{s}^2_B$). We now claim that all the other elements share the same bits in B . If not, that is, if there exists \mathbf{s}^i such that $\mathbf{s}^i_B \neq \mathbf{s}^1_B$, it must hold that $\mathbf{s}^i_\mathcal{U} = \mathbf{s}^1_\mathcal{U} \neq \mathbf{s}^2_\mathcal{U}$, which implies that $\mathbf{s}^i_B = \mathbf{s}^2_B = \mathbf{s}^1_B$, in contradiction.

We further claim that the set $S(\mathcal{T})$ must share the same \mathbf{s}_B and cannot share the same $\mathbf{s}_\mathcal{U}$. This is a simple counting argument: Since $|B| < \rho$, a set $S(\mathcal{T})$ that shares the same $\mathbf{s}_\mathcal{U}$ has size of at most $2^{|B|} \leq 2^\rho$. However, since $\mathcal{T} \in \mathcal{T}_{\text{good}}$, it holds that $|S(\mathcal{T})| > 2^\kappa$. Therefore, the set must share the same \mathbf{s}_B , and the claim follows. \blacksquare

We now show that the distinguisher cannot distinguish between the ideal and real executions, even when it asks the oracle H as (polynomially) many queries as it wishes.

First, assume that the distinguisher cannot make any queries to H . We claim that the distributions of the real and ideal executions are statistically close. Intuitively, if the adversary outputs $\mathcal{T} \in \mathcal{T}_{\text{bad}}$, then clearly the distinguisher may succeed only if the consistency check fails, which happens with probability of at most $2^{-\rho}$. On the other hand, in case where the adversary outputs $\mathcal{T} \in \mathcal{T}_{\text{good}}$, then, except for negligible probability (in ρ) it holds that $|\mathcal{U}| \geq \ell - \rho = \kappa$, and $|S(\mathcal{T})| \geq 2^\kappa$, where all the elements in $S(\mathcal{T})$ share the same bits \mathbf{s}_B . Thus, even if the distinguisher receives \mathbf{s}_B in the clear, the values $H(j, \mathbf{q}_j)$, $H(j, \mathbf{q}_j \oplus \mathbf{s})$ that are used for masking the inputs are uniformly random and independent of each other. Therefore, the simulation is indistinguishable.

Now, assume that the distinguisher can also make queries to the random oracle H . In this case, we claim that the distinguisher can distinguish only if it makes a ‘‘critical query’’. That is, for every $1 \leq j \leq m$, let $\mathbf{t}'_j = (r_j \cdot \mathbf{s}) \oplus \mathbf{t}_j \oplus (\mathbf{s} * \mathbf{e}_j)$, and recall that \mathbf{t}'_j is known to the adversary (and in fact, it can even choose $(\mathbf{t}_j, r_j, \mathbf{e}_j)$). A critical query is a query of the form $(j, \mathbf{t}'_j \oplus \mathbf{s})$ for some j , and such a query totally reveals $x_j^{\bar{r}_j}$. Clearly, conditioned on the event that the distinguisher (or

the receiver) never queries such a critical query and that $\mathbf{s}_{\mathcal{U}} \neq 0$, the distributions of the real and ideal executions are statistically close. On the other hand, the probability that they make such a query is bounded by $(t+1)/|S(\mathcal{T})| \leq (t+1) \cdot 2^{-\kappa}$, where t is the number of queries that are made by the receiver or the distinguisher. This is because as long as $\mathbf{s}_{\mathcal{U}} \neq 0$, and as long as no such a query is made, the answers of the queries are independent of the value of \mathbf{s} , and the distinguisher does not learn anything new from the queries themselves. This completes the proof. \blacksquare

We now restate Claim 3.3 and prove it. This claim is in fact, an analysis of the consistency check phase of the protocol.

Claim 3.5 (Claim 3.3, restated) *Let \mathcal{T} be as above, and let \mathcal{U} be the largest set of indices such that for every $\alpha, \beta \in \mathcal{U}$, $\mathbf{r}^\alpha = \mathbf{r}^\beta$. Assume that $|\mathcal{U}| < \ell - \rho$. Then:*

$$\Pr_{\mathbf{s}} [\text{consistent}(\mathcal{T}, \mathbf{s}) = 1] \leq 2^{-\rho}$$

and thus, $\mathcal{T} \in \mathcal{T}_{\text{bad}}$.

Proof: Let \mathcal{T} be the values that the adversary outputs, i.e., the values

$$\mathcal{T} = \{ \{ \mathbf{k}_i^0, \mathbf{k}_i^1 \}_i, \{ \mathbf{u}^i \}_i, \{ \mathcal{H}_{\alpha, \beta} \}_{\alpha, \beta} \}$$

For a pair $\alpha \in \mathcal{U}$, $\beta \in B$, we claim that the adversary passes the verification of the pair (α, β) with probability of at most $1/2$. This is because $\mathbf{r}^\alpha \neq \mathbf{r}^\beta$ and due to Lemma 3.1, if \mathcal{T} is consistent with some (s_α, s_β) then it is inconsistent with $(\overline{s}_\alpha, \overline{s}_\beta)$. Thus, there are at most 2 possible values (s_α, s_β) that are consistent with \mathcal{T} , and the adversary gets caught for the 2 other values.

We define the inconsistency graph $\Gamma = (V, E)$ as follows. The set of vertices is the set $[\ell]$. The set of edges contains all the pairs that define different \mathbf{r} 's, that is, there exists an edge (α, β) if $\mathbf{r}^\alpha \neq \mathbf{r}^\beta$. Note that since (α, β) are not consistent, the adversary gets caught in the check (α, β) with probability of at least $1/2$. We sometimes consider the complement graph (or, the consistency graph) $\overline{\Gamma} = (V, \overline{E})$. In $\overline{\Gamma}$, each edge represents that the two vertices define the same implicit input \mathbf{r} .

We now analyze the size of the set \mathcal{U} .

1. **Case 1: $\rho \leq |\mathcal{U}| < \ell - \rho$.** In this case, we have a large enough set which is consistent within itself, but is inconsistent with all the others. We claim that in this case, the adversary will get caught with probability $1 - 2^{-\rho}$.

In order to see this, consider the set $B = [\ell] \setminus \mathcal{U}$. Since $B \cup \mathcal{U} = [\ell]$, we have that $\rho < |B| \leq \ell - \rho$ as well.

Moreover, consider the inconsistency graph Γ , and remove all the edges that are between two elements of B (this can be interpreted as follows: although there is some possibility that the adversary gets caught because of these checks, we ignore them and do not consider these inconsistencies as cheating). As a result, we have a bipartite graph where the set of vertices is divided to B and \mathcal{U} . Moreover, when considering the complement graph for the resulting graph, we have two cliques, B and \mathcal{U} , and the maximal clique in this graph is at most $\ell - \rho$.

According to König's theorem [LP86], in any bipartite graph the number of edges in the maximal matching equals the minimal vertex cover. Moreover, it is easy to see that the sum of the minimum vertex cover in some graph, and the maximal clique of its complement graph equals to the overall number of vertices ℓ . We therefore conclude that the maximal matching in the graph Γ is at least ρ .

Each edge in the graph represents a check where the adversary gets caught with probability of at least $1/2$. Since there are at least ρ edges in the maximal matching in the inconsistency graph, there are at least ρ pairs for which the adversary gets caught with probability of at least $1/2$. Moreover, since we have a matching, each pair and check are independent. We therefore conclude that the adversary succeeds in its cheating with probability of at most $2^{-\rho}$, and therefore it gets caught with probability of at least $1 - 2^{-\rho}$.

2. **Case 2:** $|\mathcal{U}| < \rho$. Similarly to the previous case, we can just find a superset \mathcal{U}' that contains \mathcal{U} of size at least ρ for which we assume (artificially) that is all consistent. That is, for this set \mathcal{U}' we just ignore the checks between the elements in this set and assume that they are all consistent. After we obtain this clique (by ignoring some of the checks), we are back to the previous case.

For conclusion, we have the following: if \mathcal{T} is such that $|\mathcal{U}| < \ell - \rho$, then :

$$\Pr_{\mathbf{s}} [\text{consistent}(\mathcal{T}, \mathbf{s}) = 1] < 2^{-\rho}$$

■

3.2 Reducing the Number of Checks

In Protocol 2, in the consistency check of \mathbf{r} , we check all possible pairs $(\alpha, \beta) \in [\ell]^2$. In order to achieve higher efficiency, we want to reduce the number of checks.

Let $G = (V, E)$ be a graph for which $V = [\ell]$, and an edge (α, β) represents a check between \mathbf{r}^α and \mathbf{r}^β . In Protocol 2 the receiver asks for all possible edges in the graph (all pairs). In order to achieve better performance, we would like to reduce the number of pairs that we check. In particular, the protocol is changed so that in Step 3 of Protocol 2 the sender chooses some set of pairs (edges) $E' \subseteq V^2$, and the receiver must respond with the quadruples $\mathcal{H}_{\alpha, \beta}$ for every $(\alpha, \beta) \in E'$ that it has been asked for. The sender continues with the protocol only if all the checks have passed successfully.

Observe that after sending the values $\mathbf{u}^1, \dots, \mathbf{u}^\ell$, the sets \mathcal{U} and B (which are both subsets of $[\ell]$) are implicitly defined. In case that the set B is too large, we want to catch the adversary cheating with probability of at least $1 - 2^{-\rho}$. In order to achieve this, we should have ρ edges between B and \mathcal{U} that are pairwise non-adjacent. That is, in case the adversary defines B that is “too large”, we want to choose a set of edges E' that contains a matching between B and \mathcal{U} of size of at least ρ .

Note, however, that the sender chooses the edges E' with no knowledge whatsoever regarding the identities of \mathcal{U} and B , and thus it needs to choose a graph such that (with overwhelming probability), for any possible choice of a large B , there will be a ρ -matching between B and \mathcal{U} .

In protocol 3 we modify the consistency check of \mathbf{r} that appears in Step 3 of Protocol 2. The sender chooses for each vertex $\alpha \in [\ell]$ exactly μ out-neighbours uniformly at random. We later show that with high probability the set E' that is chosen contains a ρ -matching between the two sets B and \mathcal{U} , even for a very small value of μ (for instance, $\mu = 3$ or even $\mu = 2$).

PROTOCOL 3 (Modification for Protocol 2, Fewer Checks)

The parties run Protocol 2 with the following modifications:

Step 3 – Consistency Check of \mathbf{r} : (modified)

1. P_S chooses μ functions $\phi_0, \dots, \phi_{\mu-1}$ uniformly at random, where $\phi_i : [\ell] \rightarrow [\ell]$. It sends $\phi_0, \dots, \phi_{\mu-1}$ to the receiver P_R .
2. For every pair $\alpha \in [\ell], i \in [\mu]$, let $(\alpha, \beta) = (\alpha, \phi_i(\alpha))$. P_R defines the four values:

$$\begin{aligned} h_{\alpha,\beta}^{0,0} &= H(G(\mathbf{k}_\alpha^0) \oplus G(\mathbf{k}_\beta^0)) & h_{\alpha,\beta}^{0,1} &= H(G(\mathbf{k}_\alpha^0) \oplus G(\mathbf{k}_\beta^1)) , \\ h_{\alpha,\beta}^{1,0} &= H(G(\mathbf{k}_\alpha^1) \oplus G(\mathbf{k}_\beta^0)) & h_{\alpha,\beta}^{1,1} &= H(G(\mathbf{k}_\alpha^1) \oplus G(\mathbf{k}_\beta^1)) . \end{aligned}$$

It then sends $\mathcal{H}_{\alpha,\beta} = (h_{\alpha,\beta}^{0,0}, h_{\alpha,\beta}^{0,1}, h_{\alpha,\beta}^{1,0}, h_{\alpha,\beta}^{1,1})$ to P_S .

3. P_S checks that it receives $H_{\alpha,\phi_i(\alpha)}$ for every $\alpha \in [\ell]$ and $i \in [\mu]$. Then, for each pair $(\alpha, \beta) = (\alpha, \phi(\alpha))$ it checks that:
 - (a) $h_{\alpha,\beta}^{s_\alpha, s_\beta} = H(G(\mathbf{k}_\alpha^{s_\alpha}) \oplus G(\mathbf{k}_\beta^{s_\beta}))$.
 - (b) $h_{\alpha,\beta}^{\overline{s_\alpha}, \overline{s_\beta}} = H(G(\mathbf{k}_\alpha^{s_\alpha}) \oplus G(\mathbf{k}_\beta^{s_\beta}) \oplus \mathbf{u}^\alpha \oplus \mathbf{u}^\beta) \quad (= H(G(\mathbf{k}_\alpha^{\overline{s_\alpha}}) \oplus G(\mathbf{k}_\beta^{\overline{s_\beta}}) \oplus \mathbf{r}^\alpha \oplus \mathbf{r}^\beta))$.
 - (c) $\mathbf{u}^\alpha \neq \mathbf{u}^\beta$.

In case one of these checks fails, P_S aborts and outputs \perp .

In our modified protocol, the adversary again first outputs $\mathcal{T} = \{\{\mathbf{k}_i^0, \mathbf{k}_i^1\}_{i=1}^\ell, \mathbf{u}^1, \dots, \mathbf{u}^\ell\}$. Then, the set of checks $\Phi = \{\phi_0, \dots, \phi_{\mu-1}\}$ is chosen, and the adversary responds with $\mathcal{H} = \{\{\mathcal{H}_{\alpha,\phi_i(\alpha)}\}_{\alpha,\phi_i}\}$. We can assume that the actual secret \mathbf{s} is chosen only after \mathcal{T}, Φ and \mathcal{H} are determined. Similarly to the proof of Theorem 3.2, for a given transcript $(\mathcal{T}, \Phi, \mathcal{H})$ and a secret \mathbf{s} , we define the predicate $\text{consistent}((\mathcal{T}, \Phi, \mathcal{H}), \mathbf{s})$ that gets 1 if and only if the verification is passed for the secret \mathbf{s} (that is, that the sender does not output \perp). For a given \mathcal{T} and set of checks Φ , let $\mathcal{H}_{\mathcal{T},\Phi}$ be the set of responds that maximizes the probability to pass the verification, that is:

$$\mathcal{H}_{\mathcal{T},\Phi} \stackrel{\text{def}}{=} \text{argmax}_{\mathcal{H}} \{\Pr[\text{consistent}_{\mathbf{s}}((\mathcal{T}, \Phi, \mathcal{H}), \mathbf{s}) = 1]\} .$$

We separate all possible transcripts (\mathcal{T}, Φ) to two sets $\mathcal{T}_{\text{good}}$ and \mathcal{T}_{bad} such that:

$$\begin{aligned} \mathcal{T}_{\text{good}} &= \{(\mathcal{T}, \Phi) \mid \Pr_{\mathbf{s}}[\text{consistent}((\mathcal{T}, \Phi, \mathcal{H}_{\mathcal{T},\Phi}), \mathbf{s}) = 1] > 2^{-\rho}\} \quad \text{and} \\ \mathcal{T}_{\text{bad}} &= \{(\mathcal{T}, \Phi) \mid \Pr_{\mathbf{s}}[\text{consistent}((\mathcal{T}, \Phi, \mathcal{H}_{\mathcal{T},\Phi}), \mathbf{s}) = 1] \leq 2^{-\rho}\} . \end{aligned}$$

Observe that if a pair $(\mathcal{T}, \Phi) \in \mathcal{T}_{\text{bad}}$, then no matter what set \mathcal{H} the adversary sends, it gets caught with probability of at least $1 - 2^{-\rho}$.

The following claim is the analogue of Claim 3.3, and it bounds the size of the set B . It states that if the adversary \mathcal{A} outputs \mathcal{T} that defines $|\mathcal{U}| < \kappa$, then with probability $1 - 2^{-\rho}$ the sender will choose Φ such that $(\mathcal{T}, \Phi) \in \mathcal{T}_{\text{bad}}$.

Claim 3.6 *Let \mathcal{T} be as above, and let \mathcal{U} be the largest set of indices such that for every $\alpha, \beta \in \mathcal{U}$, $\mathbf{r}^\alpha = \mathbf{r}^\beta$. Assume that $|\mathcal{U}| < \kappa$. Then, for appropriate choice of parameters $|B|, \mu$, it holds that:*

$$\Pr_{\Phi} [(\mathcal{T}, \Phi) \in \mathcal{T}_{\text{bad}}] \geq 1 - 2^{-\rho} .$$

Proof: The partial transcript \mathcal{T} defines the two sets B and \mathcal{U} . Viewing the base-OTs $[\ell]$ as vertices in a graph, and the pairs of elements that are being checked as edges $E' = \{(\alpha, \phi_i(\alpha)) \mid \alpha \in [\ell], i \in$

$[\mu]$, we have a bipartite graph $(B \cup U, E')$ where each vertex has at least μ out edges. We want to show that with probability $1 - 2^{-\rho}$ (over the choice of Φ), there exists a ρ -matching between \mathcal{U} and B . Once there is a ρ -matching, the adversary passes the verification phase with probability of at most $2^{-\rho}$, and thus the pair (\mathcal{T}, Φ) is in \mathcal{T}_{bad} .

In order to show that in a graph there is a ρ -matching between B and \mathcal{U} , we state the following theorem which is a refinement of Hall's well-known theorem (see [LP86]). Let $N_{\mathcal{U}}(S)$ denote the set of neighbours in \mathcal{U} , for some set of vertices $S \subseteq B$, that is, $N_{\mathcal{U}}(S) = \{u \in \mathcal{U} \mid \exists v \in S, \text{ s.t. } (u, v) \in E'\}$. We have:

Theorem 3.7 *There exists a matching of size ρ between B and \mathcal{U} if and only if, for any set $S \subseteq B$, $|N_{\mathcal{U}}(S)| \geq |S| - |B| + \rho$.*

Note that we need to consider only subsets $S \subseteq B$ for which $|S| \geq |B| - \rho$ (otherwise, the condition holds trivially).

The choice of Φ is equivalent to choosing μ out edges for each vertex uniformly. We will show that for every subset of $S \subseteq B$ with $|S| \geq |B| - \rho$, it holds that $|N_{\mathcal{U}}(S)| \geq |S| - |B| + \rho$.

Let $S \subseteq B$ and $T \subseteq \mathcal{U}$. Let $X_{S,T}$ be an indicator random variable for the event that all the out-edges from S go to $B \cup T$, and all the out-edges of $\mathcal{U} \setminus T$ do not go to S (we use the term "out edges" even though the graph is not directed; our intention is simply the edges connecting these parts). As a result, $|N_{\mathcal{U}}(S)| \leq |T|$. Then, the probability that $X_{S,T}$ equals 1 is the probability that all the $\mu \cdot |S|$ out edges of S go to $B \cup T$ only, and all the $\mu \cdot (|\mathcal{U}| - |T|)$ out edges of $\mathcal{U} \setminus T$ go to $\{\ell\} \setminus S$ only. Since we have independency everywhere, we have:

$$\Pr[X_{S,T} = 1] = \left(\frac{|B| + |T|}{\ell}\right)^{|S| \cdot \mu} \cdot \left(\frac{\ell - |S|}{\ell}\right)^{(|\mathcal{U}| - |T|) \cdot \mu}$$

We are interested in the event $\sum X_{S,T}$ for all $S \subseteq B, T \subseteq \mathcal{U}$ s.t. $|B| - \rho \leq |S| \leq |B|, |T| \leq |S| - |B| + \rho$ (denote this condition by (\star)), and we want to show that it is greater than 0 with very low probability. We have:

$$\Pr \left[\sum_{S,T, \text{ s.t. } (\star)} X_{S,T} > 0 \right] \leq \sum_{S,T \text{ s.t. } (\star)} \Pr[X_{S,T} = 1] \tag{1}$$

$$\leq \sum_{S,T \text{ s.t. } (\star)} \left(\frac{|B| + |T|}{\ell}\right)^{|S| \cdot \mu} \cdot \left(\frac{\ell - |S|}{\ell}\right)^{(|\mathcal{U}| - |T|) \cdot \mu} \tag{2}$$

$$= \sum_{|S|=|B|-\rho}^{|B|} \sum_{|T|=0}^{|S|-|B|+\rho} \binom{|B|}{|S|} \cdot \binom{|\mathcal{U}|}{|T|} \cdot \left(\frac{|B| + |T|}{\ell}\right)^{|S| \cdot \mu} \cdot \left(\frac{\ell - |S|}{\ell}\right)^{(|\mathcal{U}| - |T|) \cdot \mu}$$

We do not provide an asymptotic analysis for this expression since we loose accuracy by using any upper bound for any one of the terms in it. We next compute this expression for some concrete choice of parameters. We note that the use of the union bound in Eq. (2) already reduces the tightness of our analysis, which may cause more redundant checks or base-OTs than actually needed. ■

As a result from the previous Claim, we get the following corollary:

Corollary 3.8 *Assuming that H is a random oracle, G is a pseudo-random generator, Protocol 3 with appropriate choice of parameter (ℓ, μ) security computes the $m \times OT_n$ functionality in the $\ell \times OT_\kappa$ -hybrid model in the presence of a static malicious adversary.*

Proof: The proof is based on the proof of Theorem 3.2. The simulator is the same, except for the fact that it chooses the set of checks Φ as the honest Sender in the real execution, send it to the malicious receiver and receives back the set of hashes \mathcal{H} . It then continues with the simulation as the previous proof.

Note that the verification does not pass with the same probability in the real and in the ideal executions. If $(\mathcal{T}, \Phi) \in \mathcal{T}_{\text{bad}}$, then the verification passes with probability at most $2^{-\rho}$. On the other hand, if the verification passes and $(\mathcal{T}, \Phi) \in \mathcal{T}_{\text{good}}$, then the number of consistent secrets with $(\mathcal{T}, \Phi, \mathcal{H})$ is at least 2^κ . Moreover, from Claim 3.6, $|\mathcal{U}| > \kappa$ and it holds also that $|\mathcal{U}| > |B|$. This implies that Claim 3.4 holds here as well. This implies that even if we give the distinguisher the bits \mathbf{s}_B , still there are more than 2^κ possible secrets and the simulation is indistinguishable from the same reasons as previously. \blacksquare

Concrete choice of parameters. Claim 3.6 states that the bound is achieved for an appropriate choice of parameters. We numerically computed the probability in Eq. (1) for a variety of parameters, and obtained that the probability is less than $2^{-\rho}$ with $\rho = 40$, for the following parameters:

| κ | 128 | | | | | | | 80 | | | |
|----------|-----|-----|-----|-----|-------|-------|-------|-----|-----|-----|-------|
| $ B $ | 62 | 49 | 46 | 44 | 43 | 42 | 41 | 53 | 48 | 46 | 42 |
| μ | 2 | 3 | 4 | 5 | 6 | 8 | 15 | 3 | 4 | 5 | 10 |
| ℓ | 190 | 177 | 174 | 172 | 171 | 170 | 169 | 133 | 128 | 125 | 122 |
| #-checks | 380 | 531 | 696 | 860 | 1,026 | 1,360 | 2,535 | 399 | 512 | 625 | 1,220 |

In Section 4.2, we run empirical tests to see which parameters perform best in which setting. We recall that in case we check all pairs (i.e., Protocol 2), we have either $\ell = \kappa + \rho = 128 + 40 = 168$ base-OTs with $\binom{\ell}{2} = 14,028$ checks, or $\ell = \kappa + \rho = 80 + 40 = 120$ base-OTs with 7,140 checks.

3.3 Correlation Robustness Instead of a Random Oracle

In this section, we show how a correlation robustness assumption (with respect to a high min-entropy source) suffices for proving the security of our protocol.

Correlation robust function. We first recall the standard definition of a correlation robust function from [IKNP03], as well as a stronger version of the assumption. Let U_ℓ denote the uniform distribution over strings of length ℓ .

Definition 3.9 (Correlation Robustness) *An efficiently computable function $H : \{0, 1\}^\kappa \rightarrow \{0, 1\}^n$ is correlation robust if it holds that:*

$$\{\mathbf{t}_1, \dots, \mathbf{t}_m, H(\mathbf{t}_1 \oplus \mathbf{s}), \dots, H(\mathbf{t}_m \oplus \mathbf{s})\} \stackrel{c}{\equiv} \{U_{m \cdot \kappa + m \cdot n}\}$$

where $\mathbf{t}_1, \dots, \mathbf{t}_m, \mathbf{s} \in \{0, 1\}^\kappa$ are uniformly and independently distributed. H is strongly correlation robust if for every $\mathbf{t}_1, \dots, \mathbf{t}_m \in \{0, 1\}^\kappa$ it holds that:

$$\{H(\mathbf{t}_1 \oplus \mathbf{s}), \dots, H(\mathbf{t}_m \oplus \mathbf{s})\} \stackrel{c}{\equiv} \{U_{m \cdot n}\}$$

where $\mathbf{s} \in \{0, 1\}^\kappa$ is uniform.

Another way of looking at this is as a type of *pseudorandom function*. Specifically, define $F_{\mathbf{s}}(\mathbf{t}) = H(\mathbf{t} \oplus \mathbf{s})$. Then, H is correlation robust if and only if F is a weak pseudorandom function, and H is strongly correlation robust if and only if F is a (non-adaptive) pseudorandom function. For proving the security of our protocol, we need to consider the above notions but where \mathbf{s} is chosen from a high min-entropy source. Thus, we consider the case where H is also somewhat an extractor.

Let X be a random variable taking values from $\{0, 1\}^\ell$. The min-entropy of \mathcal{X} , denoted $H_\infty(\mathcal{X})$, is: $H_\infty(\mathcal{X}) \stackrel{\text{def}}{=} \min_x \left\{ \log \frac{1}{\Pr[\mathcal{X}=x]} \right\} = -\log(\max_x \{\Pr[\mathcal{X}=x]\})$. If a source \mathcal{X} has a min entropy κ we say that \mathcal{X} is a “ κ -source”. For instance, a κ -source may be κ uniform and independent bits, together with some $\ell - \kappa$ fixed bits (in an arbitrary order), or κ uniform bits with some $\ell - \kappa$ bits that dependent arbitrarily on the first random bits. We are now ready to define min-entropy correlation robustness.

Definition 3.10 (Min-Entropy Correlation Robustness) *An efficiently computable function $H : \{0, 1\}^\ell \rightarrow \{0, 1\}^n$ is κ -min-entropy correlation robust if for all (efficiently samplable) κ -sources \mathcal{X} on $\{0, 1\}^\ell$ it holds that:*

$$\{\mathbf{t}_1, \dots, \mathbf{t}_m, H(\mathbf{t}_1 \oplus \mathbf{s}), \dots, H(\mathbf{t}_m \oplus \mathbf{s})\} \stackrel{c}{\equiv} \{U_{m \cdot \ell + m \cdot n}\}$$

where $\mathbf{t}_1, \dots, \mathbf{t}_m$ are chosen uniformly and independently at random from $\{0, 1\}^\ell$, and $\mathbf{s} \leftarrow \mathcal{X}$. H is κ -min-entropy strongly correlation robust if for all (efficiently samplable) κ -sources \mathcal{X} on $\{0, 1\}^\ell$ and every (distinct) $\mathbf{t}_1, \dots, \mathbf{t}_m \in \{0, 1\}^\ell$ it holds that:

$$\{H(\mathbf{t}_1 \oplus \mathbf{s}), \dots, H(\mathbf{t}_m \oplus \mathbf{s})\} \stackrel{c}{\equiv} \{U_{m \cdot n}\}$$

where $\mathbf{s} \leftarrow \mathcal{X}$.

In Protocol 2, the values that are used to mask the inputs of the sender are $H(\mathbf{t}_j), H(\mathbf{t}_j \oplus \mathbf{s})$ (or, $H(\mathbf{t}_j \oplus (\mathbf{s} * \mathbf{e}_j)), H(\mathbf{t}_j \oplus (\mathbf{s} * \mathbf{e}_j) \oplus \mathbf{s})$ in case the adversary uses different \mathbf{r}^i 's). Since the receiver is the one that effectively chooses the \mathbf{t}_j 's values, it may choose values that are not distributed uniformly or even choose them maliciously. As a result, we prove the security of Protocol 2 in its current form using the *strong* κ -min-entropy correlation robustness assumption.

However, it is also possible to modify the protocol and rely only on κ -min-entropy correlation robustness, as follows. In Step 4c (of Protocol 2), in each iteration $1 \leq j \leq m$, the sender chooses a random value $\mathbf{d}_j \in \{0, 1\}^\ell$, and sends the values $(\mathbf{d}_j, y_j^0, y_j^1)$, where:

$$y_j^0 = x_j^0 \oplus H(j, \mathbf{q}_j \oplus \mathbf{d}_j) \quad \text{and} \quad y_j^1 = x_j^1 \oplus H(j, \mathbf{q}_j \oplus \mathbf{d}_j \oplus \mathbf{s}).$$

Then, P_R computes $x_j = y_j^{r_j} \oplus H(j, \mathbf{t}_j \oplus \mathbf{d}_j)$. Since the \mathbf{d}_j values are chosen last, this ensures that the values used inside H are always uniformly distributed. Thus, κ -min-entropy correlation robustness suffices.

In Step 3 of Protocol 2 we also use the function H ; however, the properties needed from H for these invocations are collision resistance (for the case of a corrupted receiver) and κ -min entropy correlation robustness (for the case of a corrupted sender). In order to emphasize the differences between the function used for the verification and the function used for the transfer phase, we denote the former by h (i.e., the one used in Step 3 of the protocol), and the latter by H (i.e., the one used in Step 4c).

Theorem 3.11

1. Assume that H is strongly κ -min-entropy correlation robust, h is a collision resistant function and κ -min-entropy correlation robust, and G is a pseudo-random generator. Then, Protocol 2 securely computes the $m \times OT_n$ functionality in the $\ell \times OT_\kappa$ -hybrid model in the presence of a static malicious adversary.
2. Assume that H is κ -min-entropy correlation robust, h is a collision resistant function and κ -min-entropy correlation robust, and G is a pseudo-random generator. Then, the above-described modified protocol securely computes the $m \times OT_n$ functionality in the $\ell \times OT_\kappa$ -hybrid model in the presence of a static malicious adversary.

Proof: We prove the first item in the theorem. The second is proven in almost the same way. Moreover, we consider for now the original protocol (i.e., Protocol 2, where the checks of all pairs are performed). We later show how to consider the protocol with the reduced number of checks.

We conclude security for the corrupted sender as in Claim C.1 in Appendix C.1.

Recall that in both the ideal and real executions, the outputs of the execution consist of the randomness of the adversary, its view (the messages it receives during the execution) and the output of the honest party. The randomness of the adversary uniquely defines the messages it sends in the first round $\mathcal{T} = \{\mathbf{k}_i^0, \mathbf{k}_i^1\}_{i=1}^\ell, \mathbf{u}^1, \dots, \mathbf{u}^\ell, \mathcal{H}_{\alpha, \beta}\}$. The view of the adversary consists of the messages it receives in the last round of the protocol, that is, $\{y_i^0, y_i^1\}_{i=1}^m$. We now consider two cases; the first in which the randomness of the adversary defines \mathcal{T} for which $\mathcal{T} \in \mathcal{T}_{\text{bad}}$ and the second case where $\mathcal{T} \in \mathcal{T}_{\text{good}}$.

A outputs $\mathcal{T} \in \mathcal{T}_{\text{bad}}$. In such a case, in both executions the adversary gets caught with probability $1 - 2^{-\rho}$. This is because both the simulator in the ideal execution and the honest sender in the real execution choose a secret \mathbf{s} uniformly at random in $\{0, 1\}^\ell$, and it holds that:

$$\Pr[\text{consistent}(\mathcal{T}, \mathbf{s}) = 1] \leq 2^{-\rho}$$

In case the verification does not pass, both the simulator in the ideal execution and the honest sender in the real execution halt the execution immediately and do not transfer the values $\{y_i^0, y_i^1\}_{i=1}^\ell$. As a result, the two execution are clearly identical, since the adversary has no view and the output of the honest party in both cases are \perp . The only possibility of failure is in case where the verification passes although $\mathcal{T} \in \mathcal{T}_{\text{bad}}$, which happens with probability $2^{-\rho}$.

A outputs $\mathcal{T} \in \mathcal{T}_{\text{good}}$. Even though $\mathcal{T} \in \mathcal{T}_{\text{good}}$, there is still a noticeable probability that the verification will not pass. Since the secret \mathbf{s} is chosen exactly the same way in both executions, the verification passes or fails with the exact same probability.

If the verification does not pass, i.e., $\mathbf{s} \notin \mathcal{S}(\mathcal{T})$, then in both the real and the ideal executions there is no transmission, and therefore the both executions are identical as above.

We left with the case where $\mathcal{T} \in \mathcal{T}_{\text{good}}$ and that $\mathbf{s} \in \mathcal{S}(\mathcal{T})$. In such a case, there is a transmission in both executions. We show that the two are indistinguishable by a mental experiment and consider the following three executions:

1. The real execution, conditioned on the event where $\mathcal{T} \in \mathcal{T}_{\text{good}}$ and $\mathbf{s} \in \mathcal{S}(\mathcal{T})$.
2. The real execution, conditioned on the event that $\mathcal{T} \in \mathcal{T}_{\text{good}}$ and \mathbf{s} is chosen from the κ -source $\mathcal{X}(\mathcal{T})$. Below, in Claim 3.12 we show how one can sample from the set $\mathcal{S}(\mathcal{T}) = \{\mathbf{s} \in \{0, 1\}^\ell \mid \text{consistent}(\mathcal{T}, \mathbf{s}) = 1\}$ efficiently.

3. The ideal execution, conditioned on the event that $\mathcal{T} \in \mathcal{T}_{\text{good}}$ and $\mathbf{s} \in \mathcal{S}(\mathcal{T})$.

The only difference between execution 2 and execution 3 are the values $\{y_j^{1-r_j}\}_{j=1}^m$. We recall that in the ideal execution, these values are uniform and independent, whereas in the real execution for $j = 1, \dots, m$, it holds that $y_j^{1-r_j} = x_j^{1-r_j} \oplus H(\mathbf{t}'_j \oplus \mathbf{s})$. However, from the fact that H is a strongly κ -min entropy correlation robust (as in Definition 3.10), executions 2 and 3 are computationally-indistinguishable.

The only difference between execution 1 and 2 is the way \mathbf{s} is chosen. In the real execution, we condition on the case where $\mathbf{s} \in \mathcal{S}(\mathcal{T})$, and thus \mathbf{s} is distributed uniformly in $\mathcal{S}(\mathcal{T})$. In execution 2, \mathbf{s} is chosen uniformly from the set $\mathcal{S}(\mathcal{T})$. These two executions are distributed identically.

The case of Protocol 3. We now consider the protocol with the fewer number of checks. Again, the messages $\mathcal{T} = \{\{\mathbf{k}_i^0, \mathbf{k}_i^1\}_{i=1}^\ell, \mathbf{u}^1, \dots, \mathbf{u}^\ell\}$ depends only on the randomness of \mathcal{A} and therefore are the same in both executions. Both the honest sender in the real execution, and the simulator in the ideal execution, choose the functions Φ with the same distribution, and therefore the hashes $\mathcal{H} = \{\mathcal{H}_{\alpha, \phi(\alpha)}\}_{\alpha \in [\ell], \phi \in \Phi}$ have the same distribution. As the previous protocol, the case where $(\mathcal{T}, \Phi) \in \mathcal{T}_{\text{bad}}$ happens with the same probability in both executions and the view of the adversary is the same in both executions.

Given $(\mathcal{T}, \Phi, \mathcal{H})$, the case of $(\mathcal{T}, \Phi) \in \mathcal{T}_{\text{good}}$, but for which $\mathbf{s} \notin \mathcal{S}(\mathcal{T}, \Phi, \mathcal{H})$, where $\mathcal{S}(\mathcal{T}, \Phi, \mathcal{H}) = \{\mathbf{s} \in \{0, 1\}^\ell \mid \text{consistent}((\mathcal{T}, \Phi, \mathcal{H}), \mathbf{s}) = 1\}$ also occurs with the same probability, and the view of the adversary and the output of the honest party are clearly the same in both execution.

The case where $(\mathcal{T}, \Phi) \in \mathcal{T}_{\text{good}}$ and $\mathbf{s} \in \mathcal{S}(\mathcal{T}, \Phi, \mathcal{H})$ is handled as in the equivalent case above. Specifically, consider an execution where \mathbf{s} is chosen from the source $\mathcal{X}(\mathcal{T}, \Phi, \mathcal{H})$ is defined below. This execution is identical to the real, and by the correlation robustness property of H , this execution is indistinguishable from the real, since H is a strongly κ -min entropy correlation robust. ■

Claim 3.12 *For any given transcript $\mathcal{T} \in \mathcal{T}_{\text{good}}$, there exists an efficient procedure that samples a uniform secret \mathbf{s} from $\mathcal{S}(\mathcal{T})$. This procedure is a κ -source.*

Proof: We want to show that given the transcript that the adversary has outputted, we can extract the constraints that are defined by these values, and the bits that are learned from the fact that the verification has passes. This will give us the ability to sample a value from $\mathcal{S}(\mathcal{T})$. Note that just sampling a random $\mathbf{s} \in \{0, 1\}^\ell$, and performing the same checks as in the honest execution is not enough, since there are $\{0, 1\}^\ell$ possible secrets overall, whereas $|\mathcal{S}(\mathcal{T})|$ may be an order of 2^κ . As a result, the probability that a random $\mathbf{s} \in \{0, 1\}^\ell$ is a consistent secret may be too small.

For a pair (α, β) , consider $H_{\alpha, \beta} = (h_{\alpha, \beta}^{0,0}, h_{\alpha, \beta}^{0,1}, h_{\alpha, \beta}^{1,0}, h_{\alpha, \beta}^{1,1})$. Let $\text{correct}_{p,q}(\mathcal{H}_{\alpha, \beta}) \in \{0, 1\}^4$ be a predicate that its value is 1 if and only if $h_{\alpha, \beta}^{p,q} = H(G(\mathbf{k}_\alpha^p) \oplus G(\mathbf{k}_\beta^q))$. Finally, let

$$\text{correct}(\mathcal{H}_{\alpha, \beta}) \stackrel{\text{def}}{=} (\text{correct}_{0,0}(\mathcal{H}_{\alpha, \beta}), \text{correct}_{0,1}(\mathcal{H}_{\alpha, \beta}), \text{correct}_{1,0}(\mathcal{H}_{\alpha, \beta}), \text{correct}_{1,1}(\mathcal{H}_{\alpha, \beta})) .$$

We also assume that in cases where $\mathbf{r}^\alpha \neq \mathbf{r}^\beta$, whenever the adversary sets $h_{\alpha, \beta}^{p,q}$ that is incorrect, it sets its value to be $H(G(\mathbf{k}_\alpha^{\bar{p}}) \oplus G(\mathbf{k}_\beta^{\bar{q}}) \oplus \mathbf{u}^\alpha \oplus \mathbf{u}^\beta)$ in order to maximizes the success probability of the verification. We note that this condition can be verified as well, and generate new constraints in case it does not hold. Algorithm 1 describes how one can sample a consistent secret \mathbf{s} from a given transcript \mathcal{T} .

ALGORITHM 1 (The κ -source $\mathcal{X}(\mathcal{T})$: sampling a uniform \mathbf{s} from $\mathcal{S}(\mathcal{T})$)**Input:** A transcript $\mathcal{T} = \{\{\mathbf{k}_i^0, \mathbf{k}_i^1\}_{i=1}^\ell, \mathbf{u}^1, \dots, \mathbf{u}^\ell, \mathcal{H}_{\alpha,\beta}\}$.

1. Extract the following constraints regarding \mathbf{s} and store them. For every pair (α, β) extract $\mathbf{r}^\alpha = \mathbf{u}^\alpha \oplus G(\mathbf{k}_\alpha^0) \oplus G(\mathbf{k}_\alpha^1)$ and \mathbf{r}^β in a similar manner. Then:
 - (a) If $\mathbf{r}^\alpha = \mathbf{r}^\beta$:
 - i. If $\text{correct}(\mathcal{H}_{\alpha,\beta}) = (1, 1, 1, 1)$ then no new constraint is added.
 - ii. If $\text{correct}(\mathcal{H}_{\alpha,\beta}) = (1, 0, 0, 1)$ then add the constraint $s_\alpha \oplus s_\beta = 0$.
 - iii. If $\text{correct}(\mathcal{H}_{\alpha,\beta}) = (0, 1, 1, 0)$ then add the constraint $s_\alpha \oplus s_\beta = 1$.
 - (b) If $\mathbf{r}^\alpha \neq \mathbf{r}^\beta$:
 - i. If $\text{correct}(\mathcal{H}_{\alpha,\beta}) = (1, 1, 0, 0)$ then add the constraint $s_\alpha = 0$.
 - ii. If $\text{correct}(\mathcal{H}_{\alpha,\beta}) = (0, 0, 1, 1)$ then add the constraint $s_\alpha = 1$.
 - iii. If $\text{correct}(\mathcal{H}_{\alpha,\beta}) = (1, 0, 1, 0)$ then add the constraint $s_\beta = 0$.
 - iv. If $\text{correct}(\mathcal{H}_{\alpha,\beta}) = (0, 1, 0, 1)$ then add the constraint $s_\beta = 1$.
 - (c) If $\mathbf{r}^\alpha \neq \mathbf{r}^\beta$ and only one position of $\mathcal{H}_{\alpha,\beta}$ is correct - learn both (s_α, s_β) and add this as a constraint. (e.g., if $\text{correct}(\mathcal{H}_{\alpha,\beta}) = (1, 0, 0, 0)$ then add the constraint $(s_\alpha, s_\beta) = (0, 0)$.)
2. If some of the constraints are contradicting, abort and output \perp .
3. Otherwise, choose a random $\mathbf{s} \in \{0, 1\}^\ell$ under the constraints that were stored above, and output it.

It is easy to see that the possible outputs of the algorithm are exactly the set $\mathcal{S}(\mathcal{T})$. Moreover, since $\mathcal{T} \in \mathcal{T}_{\text{good}}$, it holds that $|\mathcal{S}(\mathcal{T})| \geq 2^\kappa$. As a result, for every possible output \mathbf{s} of the algorithm $\mathcal{X}(\mathcal{T})$, it holds that $\Pr[\mathcal{X}(\mathcal{T}) = \mathbf{s}] \leq 2^{-\kappa}$, and thus the min-entropy of $\mathcal{X}(\mathcal{T})$ is κ . \blacksquare

Algorithm 1 was designed for the variant of the protocol where we check all pairs. An equivalent source $\mathcal{X}(\mathcal{K}, \Phi, \mathcal{H})$ for the variant of the protocol that does not check all pairs, can also be constructed in a similar manner. The only difference between the two algorithms is that we do not run over all possible pairs (α, β) in Step 1 of the algorithm, but rather only all pairs $(\alpha, \phi(\alpha))_{\phi \in \Phi}$. This is a κ -source for every $(\mathcal{T}, \Phi) \in \mathcal{T}_{\text{good}}$, since the number of possible outputs \mathcal{S} is at least 2^κ .

3.4 Achieving Covert Security

In this section, we present a more efficient protocol (with fewer base-OTs and checks) with the property that any deviation from the protocol that can result in a breach of security will be detected with probability at least $1/2$. For details on the definition of covert security, we refer to [AL10]. Our protocol below is secure under the *strong explicit-cheat formulation* with deterrent factor $\epsilon = \frac{1}{2}$.

As in the malicious case, given the set of keys $\{\mathbf{k}_i^0, \mathbf{k}_i^1\}$, and the messages u^1, \dots, u^ℓ , the sets B and \mathcal{U} are implicitly defined, and we want to catch the adversary if its behavior defines a set B with “high” cardinality. Here, in contrast to the malicious case, we will be content with catching the adversary with probability $1/2$, instead of $1 - 2^{-\rho}$ as in the case of malicious adversaries. As we will show below, our approach for the consistency check of \mathbf{r} enables us to achieve a deterrent factor of $1/2$ at the cost of very few consistency checks. Concretely, it will be enough to use 7 checks of pairs only.

The protocol. In Step 3 of Protocol 2, the sender chooses t random pairs $\{(\alpha_i, \beta_i)\}_{i=1}^t$ uniformly and independently at random, and sends them to the receiver. The receiver sends $\mathcal{H}_{\alpha_i, \beta_i}$ for each

pair (α_i, β_i) that it was asked. Then, the sender performs the same checks as in the previous protocol: It checks that the receiver replied with hashes for all the pairs (α_i, β_i) that it was asked for, and that the hashes that were sent are correct (i.e., as in Step 3b of Protocol 2).

The analysis. Although at first sight the analysis below ignores attacks of Type 2, these attacks are still taken into consideration. This is because whenever the adversary tries to cheat and learn bits of \mathbf{s} where $\mathbf{r}^\alpha = \mathbf{r}^\beta$, it gets caught doing so with probability $1/2$, which is exactly the deterrent factor. The analysis therefore focuses on the case that the adversary cheats when $|B|$ is “too large”, and shows that when we have t checks and $|B|$ is large enough, then the probability that the adversary passes the verification is less than $1/2$.

We again consider the graph of checks, and let $V = [\ell]$ and the edges are all possible checks. We divide $[\ell]$ to B and \mathcal{U} , and we show that when using t checks, the probability that the adversary succeeds to pass the verification when B is “large” is less than $1/2$.

There are ℓ^2 edges overall, where $2|B| \cdot |\mathcal{U}|$ are edges between B and \mathcal{U} , and $|B|^2 + |\mathcal{U}|^2$ edges are between B and B , or \mathcal{U} and \mathcal{U} . We say that an edge is “good” if it goes between B and \mathcal{U} . Recall that in such a check, the adversary is caught with probability at least $1/2$.

For the first edge that is chosen, the probability that it is a good edge is $2|B| \cdot |\mathcal{U}|/\ell^2$. However, once this specific edge between B and \mathcal{U} is chosen, an edge between B and \mathcal{U} that is pairwise non-adjacent with the previously chosen edge is not longer good, since the probability that the adversary will get caught here is not $1/2$. Therefore, we denote by good_i the probability of choosing the $(i+1)$ th “good” edge. That is, the probability that edge e_j is good, conditioned on the event that i good edges were previously chosen in the set $\{e_1, \dots, e_{j-1}\}$. We have that:

$$\text{good}_i = \frac{2 \cdot (|B| - i) \cdot (|\mathcal{U}| - i)}{\ell^2}.$$

This holds because once a good edge is chosen, we do not want to choose an edge that is adjacent to it. As a result, with each good edge that is chosen, the effective size of the set B and \mathcal{U} is decreased by 1.

In contrast, we denote by bad_i the probability that the next chosen edge is bad, given that there were i previous good edges. That is, a bad edge is either an edge between B and B , an edge between \mathcal{U} and \mathcal{U} , or is adjacent to one of the $2i$ vertices of the previously chosen good edges. This probability is as follows:

$$\text{bad}_i = \frac{|B|^2 + |\mathcal{U}|^2 + 2i \cdot |\mathcal{U}| + 2i \cdot |B| - 2i^2}{\ell^2} = \frac{|B|^2 + |\mathcal{U}|^2 + 2i(\ell - i)}{\ell^2}$$

That is, a bad edge can be either an edge from B to B , \mathcal{U} to \mathcal{U} , or an edge between the i vertices that were chosen with any other vertex. Note, however, that there are some edges that are counted twice and thus we remove $2i^2$. In addition, observe that $\text{good}_i + \text{bad}_i = 1$.

When we have t checks, we may have between 0 to t good edges. In case there are d good edges, the probability that the adversary succeeds to cheat is 2^{-d} . In order to ease the calculation, let good be the maximal probability of $\text{good}_0, \dots, \text{good}_{t-1}$, and let bad be the maximal probability of $\text{bad}_0, \dots, \text{bad}_t$. We get that:

$$\text{good} = \frac{2 \cdot |B| \cdot |\mathcal{U}|}{\ell^2}$$

and for $t < \ell/2$:

$$\text{bad} = \frac{|B|^2 + |\mathcal{U}|^2 + 2t(\ell - t)}{\ell^2}.$$

Now, consider the edges e_1, \dots, e_t . The probability that the adversary succeeds in its cheating is the union of succeeds in cheating in each possible combination of checks. In particular, we may have $d = 0, \dots, t$ good edges, and for each d , there are $\binom{t}{d}$ possible ways to order d good edges and $t - d$ “bad” edges. Finally, when we have d good edges, the probability that the adversary succeeds to cheat is 2^{-d} . We therefore have that the probability that the adversary successfully cheats without being caught is less than:

$$\sum_{d=0}^t \binom{t}{d} \cdot \text{good}^d \cdot \text{bad}^{t-d} \cdot 2^{-d} = \sum_{d=0}^t \binom{t}{d} \cdot \left(\frac{1}{2} \cdot \text{good}\right)^d \cdot \text{bad}^{t-d} = \left(\frac{1}{2} \cdot \text{good} + \text{bad}\right)^t.$$

It is easy to verify that this probability is less than 0.5 for $|B| = 38$, $|\mathcal{U}| = 128$ (and so overall $\ell = 166$), with only 7 checks. In which case, we have that $\text{good} = 0.353$, $\text{bad} = 0.728$, and the probability is less than 0.495.

4 Performance Evaluation

We experimentally compare the performance of our protocols to previous works using the same programming language and running benchmarks on the same machines: We first describe our implementation (§4.1), empirically evaluate and compare the identified active and covert parameters of §3.2 and §3.4 (§4.2), and compare our work to the active-secure protocol of [NNOB12] with optimizations of [FN13] and to the passive-secure protocol of [IKNP03] with optimizations from [ALSZ13] (§4.3).

Benchmarking Environment: We run our experiments in two settings: a local setting and a cloud setting. In the *local setting*, the sender and receiver routines run on two Desktop PCs which each have 16 GB RAM, an Intel Haswell i7-4770K CPU with 4 cores and AES-NI support, and are connected via Gigabit Ethernet. In the *cloud setting*, we run the OT sender routine on an Amazon EC2 m3.medium instance with a 2.5 GHz, Intel Xeon E5-2670v2 CPU and 3.75 memory located in North Virginia (US East) and run the OT receiver routine on one of our Desktop PCs in Europe. The average bandwidth usage in the cloud setting was 52 MBit/s and the average ping latency (round-trip-time) was 95 ms.

4.1 Implementation

We build on the passive-secure and publicly available OT extension C++ implementation of [ALSZ13]. We perform the OT extension protocol and consistency checks block-wise, i.e., we split m OTs into b blocks of size $w = 2^{18}$, with $b = \lceil \frac{m}{w} \rceil$. These blocks can be processed independently of each other and using multiple threads. For all experiments we evaluate the random OT version of [ALSZ13], since the additional overhead to obtain the traditional OT functionality is equal for all protocols, and output $n = 8$ -bit strings. For the base-OTs we use [NP01] for the passive-secure OT extension protocol and [PVW08] in decryption mode with security based on the Decisional Diffie-Hellman (DDH) assumption for the covert- and active-secure OT extension protocols; we implement both using elliptic curves. We assume $\kappa = 128$ -bit long-term security with $\rho = 40$ statistical security. Further implementation details are given in Appendix §A.

4.2 Parameter Evaluation

We evaluate the asymptotic communication and run-time in the local and cloud setting on 2^{23} random OTs for our most promising active security (cf. Table 3.2) and covert security (cf. §3.4) parameters, and compare them to the active-secure protocol of [NNOB12] with $\ell = \lceil \frac{8}{3}\kappa \rceil = 342$ base-OTs and $\ell/2 = 171$ checks, and to the passive-secure protocol of [IKNP03] with $\ell = 128$ base-OTs and no checks. The results are depicted in Table 2 where the parameters are given as (#base-OTs;#checks). We also include the pairwise comparison Protocol 2 (which performs all possible checks) with parameters (168;14,028) and discuss its special features in Appendix §A.3.

| Security | Active | | | | | | Covert | Passive |
|----------------------|----------|---------------|---------|---------------|------------|------------|---------------|---------------|
| Parameters | [NNOB12] | 190;380 | 177;531 | 174;696 | 170;1,360 | 168;14,028 | 166;7 | [IKNP03] |
| Comm. [MB] | 342 | 191 | 178 | 175 | 173 | 195 | 166 | 128 |
| <i>Local Setting</i> | | | | | | | | |
| Run-time [s] | 16.988 | 11.938 | 13.201 | 18.218 | 25.918 | 221.382 | 10.675 | 9.579 |
| <i>Cloud Setting</i> | | | | | | | | |
| Run-time [s] | 110.223 | 64.698 | 63.845 | 63.712 | 83.414 | 454.595 | 46.718 | 33.838 |

Table 2: Run-time and communication for active, covert, and passive security using different parameters (#base-OTs;#checks) on 2^{23} random OTs. Minimum values are marked in bold.

For the communication we can observe that our parameter sets have 50% – 55% of the communication of [NNOB12]. Furthermore, while decreasing the number of base-OTs reduces the overall communication until 170 base-OTs, the overhead in communication for sending the consistency check hashes outweighs the gains from the reduced number of base-OTs. Hence, using less than 170 base-OTs for block-size $w = 2^{18}$ would increase both communication and computation complexity.

For the run-time we can observe that our best-performing parameter has 70% of the run-time of [NNOB12] in the local setting and 58% of the run-time in the cloud setting. Furthermore, the best-performing parameter differs between the local and cloud setting: while the (190;380) parameter performs best in the local setting, the (174;696) parameter achieves the lowest run-time in the cloud setting. This can be explained by the smaller bandwidth of the cloud setting, which influences the run-time of all parameters differently. For instance, when switching from the local to the cloud setting, the run-time of [NNOB12] increases by factor 6.5, whereas that of our pairwise comparison Protocol 2 with parameter (168;14,028) only increases by factor 2. As expected, the covert parameter (166;7) performs better than the parameters for active security.

4.3 Comparison with Related Work

We empirically evaluate and compare our protocol on a varying number of OTs in its active and covert versions to the passive-secure OT extension protocol of [IKNP03] with optimizations of [KK13, ALSZ13], and the active-secure OT extension protocol of [NNOB12] with optimizations of [FN13]. The results for the local and cloud setting are given in Figure 1. We benchmark the protocols on an exponentially increasing number of OTs: from 2^{10} to 2^{29} for the local setting and from 2^{10} to 2^{26} for the cloud setting. The passive-secure [IKNP03] serves as bottom-line for the performance of the other protocols to show the (small) gap to the covert- and active-secure protocols. For our protocol we use the parameters from our parameter evaluation in §4.2 which were shown to perform best in the respective setting, i.e., (190;380) for the local setting, (174;696) for the cloud setting, and (166;7) for covert security. For the [NNOB12] protocol we use $\ell = \lceil \frac{8}{3}\kappa \rceil = 342$ base-OTs

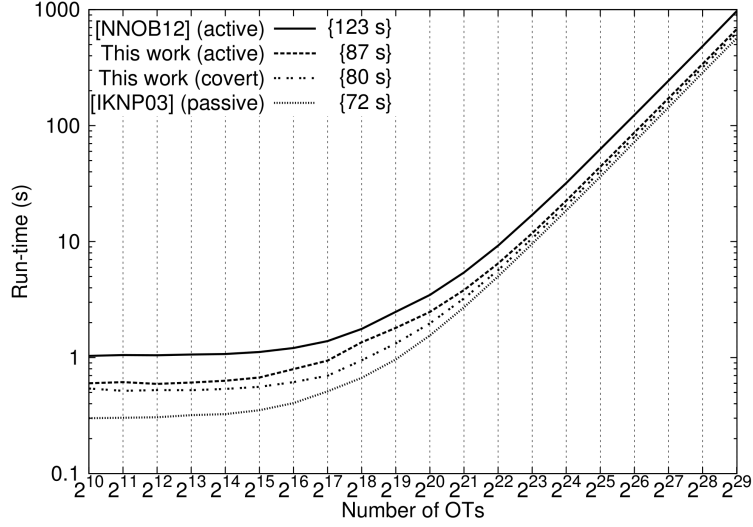
and $\ell/2 = 171$ checks. We excluded the active-secure protocol of [Lar14], since its communication overhead is at least two orders of magnitude higher than for the evaluated protocols and simply transferring the required data would result in higher run-times than those of the other protocols.

For the results in the local setting we can observe that our active-secure OT extension protocol outperforms the [NNOB12] protocol for all OTs tested on and scales better with increasing number of OTs. Furthermore, our active-secure protocol converges towards the passive-secure [IKNP03] protocol when more OTs are performed, decreasing the overhead for active security down to 121% for 2^{26} OTs, compared to an overhead of 171% for the [NNOB12] protocol. The convergence of our protocol can be explained by the amortizing costs of the consistency checks. Since the consistency checks are performed on blocks of fixed width 2^{18} , their amortization happens for a larger number of OTs. The covert version of our protocol has only 111% overhead compared to the passive-secure protocol.

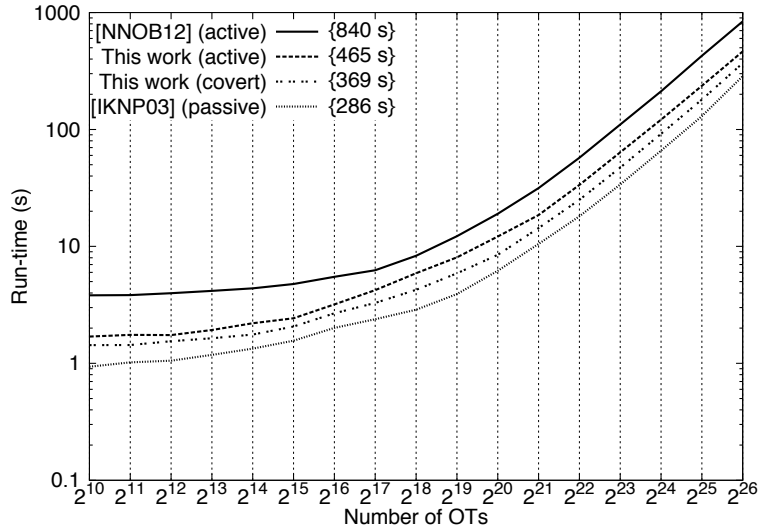
In the cloud setting, the performance of all protocols decreases, as expected. However, the performance of the passive-secure protocol decreases less significantly compared to the covert- and active-secure protocols. This can be explained by the smaller communication complexity of the passive-secure protocol, since the run-time overhead scales with the communication overhead of the respective protocol. For the active-secure protocol of [NNOB12] with communication overhead of 267% compared to the passive-secure protocol, the run-time overhead increases from 171% to 294%. In comparison, for our active-secure protocol with communication overhead of 136%, the run-time overhead increases from 121% to 163%. Finally, for our covert protocol with communication overhead of 129%, the run-time overhead increases from 111% to 129%.

Acknowledgments

We thank Peter Scholl for pointing out a flaw in security of corrupted sender in an earlier version of this paper.



(a) Local Setting



(b) Cloud Setting

Figure 1: Run-time for random OT extension protocols for 8-bit strings with active, covert, and passive security in the local- and cloud setting. Time for 2^{26} OTs given in {}.

References

- [AL10] Y. Aumann and Y. Lindell. Security against covert adversaries: Efficient protocols for realistic adversaries. *Journal of Cryptology*, 23(2):281–343, 2010.
- [ALSZ13] G. Asharov, Y. Lindell, T. Schneider, and M. Zohner. More efficient oblivious transfer and extensions for faster secure computation. In *ACM Computer and Communications Security (CCS'13)*, pages 535–548. ACM, 2013. Code: <http://crypto.de/code/OTExtension>.

- [Bea96] D. Beaver. Correlated pseudorandomness and the complexity of private computations. In *Symposium on the Theory of Computing (STOC'96)*, pages 479–488. ACM, 1996.
- [DCW13] C. Dong, L. Chen, and Z. Wen. When private set intersection meets big data: An efficient and scalable protocol. In *ACM Computer and Communications Security (CCS'13)*, pages 789–800. ACM, 2013.
- [DLT14] I. Damgård, R. Lauritsen, and T. Toft. An empirical study and some improvements of the MiniMac protocol for secure computation. In *Security and Cryptography for Networks (SCN'14)*, volume 8642 of *LNCS*, pages 398–415. Springer, 2014.
- [DZ13] I. Damgård and S. Zakarias. Constant-overhead secure computation of Boolean circuits using preprocessing. In *Theory of Cryptography Conference (TCC'13)*, volume 7785 of *LNCS*, pages 621–641. Springer, 2013.
- [EGL85] S. Even, O. Goldreich, and A. Lempel. A randomized protocol for signing contracts. *Communications of the ACM*, 28(6):637–647, 1985.
- [FJN14] T. K. Frederiksen, T. P. Jakobsen, and J. B. Nielsen. Faster maliciously secure two-party computation using the GPU. In *Security and Cryptography for Networks (SCN'14)*, volume 8642 of *LNCS*, pages 358–379. Springer, 2014.
- [FN13] T. K. Frederiksen and J. B. Nielsen. Fast and maliciously secure two-party computation using the GPU. In *Applied Cryptography and Network Security (ACNS'13)*, volume 7954 of *LNCS*, pages 339–356. Springer, 2013.
- [GMW87] O. Goldreich, S. Micali, and A. Wigderson. How to play any mental game or a completeness theorem for protocols with honest majority. In *Symposium on Theory of Computing (STOC'87)*, pages 218–229. ACM, 1987.
- [HEK12] Y. Huang, D. Evans, and J. Katz. Private set intersection: Are garbled circuits better than custom protocols? In *Network and Distributed System Security (NDSS'12)*. The Internet Society, 2012.
- [HIKN08] D. Harnik, Y. Ishai, E. Kushilevitz, and J. Buus Nielsen. OT-combiners via secure computation. In *Theory of Cryptography Conference (TCC'08)*, volume 4948 of *LNCS*, pages 393–411. Springer, 2008.
- [HKK⁺14] Y. Huang, J. Katz, V. Kolesnikov, R. Kumaresan, and A. J. Malozemoff. Amortizing garbled circuits. In *Advances in Cryptology – CRYPTO'14*, volume 8617 of *LNCS*, pages 458–475. Springer, 2014.
- [IKNP03] Y. Ishai, J. Kilian, K. Nissim, and E. Petrank. Extending oblivious transfers efficiently. In *Advances in Cryptology – CRYPTO'03*, volume 2729 of *LNCS*, pages 145–161. Springer, 2003.
- [IR88] R. Impagliazzo and S. Rudich. Limits on the provable consequences of one-way permutations. In *Advances in Cryptology (CRYPTO'88)*, volume 403 of *LNCS*, pages 8–26. Springer, 1988.

- [JKO13] M. Jawurek, F. Kerschbaum, and C. Orlandi. Zero-knowledge using garbled circuits: How to prove non-algebraic statements efficiently. In *ACM Computer and Communications Security (CCS'13)*, pages 955–966. ACM, 2013.
- [KK13] V. Kolesnikov and R. Kumaresan. Improved OT extension for transferring short secrets. In *Advances in Cryptology – CRYPTO'13*, volume 8043 of *LNCS*, pages 54–70. Springer, 2013.
- [KSS12] B. Kreuter, A. Shelat, and C. Shen. Billion-gate secure computation with malicious adversaries. In *USENIX Security Symposium'12*, pages 285–300. USENIX, 2012.
- [Lar14] E. Larraia. Extending oblivious transfer efficiently, or - how to get active security with constant cryptographic overhead. In *Progress in Cryptology – LATINCRYPT'14*, *LNCS*. Springer, 2014. To appear. Online: <http://eprint.iacr.org/2014/692>.
- [LOS14] E. Larraia, E. Orsini, and N. P. Smart. Dishonest majority multi-party computation for binary circuits. In *Advances in Cryptology – CRYPTO'14*, volume 8617 of *LNCS*, pages 495–512. Springer, 2014.
- [LP86] L. Lovász and M.D. Plummer. *Matching Theory*. Akadémiai Kiadó, Budapest, 1986. Also published as Vol. 121 of the North-Holland Mathematics Studies, North-Holland Publishing, Amsterdam.
- [LP07] Y. Lindell and B. Pinkas. An efficient protocol for secure two-party computation in the presence of malicious adversaries. In *Advances in Cryptology – EUROCRYPT'07*, volume 4515 of *LNCS*, pages 52–78. Springer, 2007.
- [LP11] Y. Lindell and B. Pinkas. Secure two-party computation via cut-and-choose oblivious transfer. In *Theory of Cryptography Conference (TCC'11)*, volume 6597 of *LNCS*, pages 329–346. Springer, 2011.
- [LPS08] Y. Lindell, B. Pinkas, and N. P. Smart. Implementing two-party computation efficiently with security against malicious adversaries. In *Security and Cryptography for Networks (SCN'08)*, volume 5229 of *LNCS*, pages 2–20. Springer, 2008.
- [LR14] Y. Lindell and B. Riva. Cut-and-choose Yao-based secure computation in the online/offline and batch settings. In *Advances in Cryptology – CRYPTO'14*, volume 8617 of *LNCS*, pages 476–494. Springer, 2014.
- [LZ13] Y. Lindell and H. Zarosim. On the feasibility of extending oblivious transfer. In *Theory of Cryptography Conference (TCC'13)*, volume 7785 of *LNCS*, pages 519–538. Springer, 2013.
- [Nie07] J. B. Nielsen. Extending oblivious transfers efficiently - how to get robustness almost for free. *IACR Cryptology ePrint Archive*, 2007:215, 2007. Online: <http://eprint.iacr.org/2007/215>.
- [NNOB12] J. B. Nielsen, P. S. Nordholt, C. Orlandi, and S. S. Burra. A new approach to practical active-secure two-party computation. In *Advances in Cryptology – CRYPTO'12*, volume 7417 of *LNCS*, pages 681–700. Springer, 2012.

- [NP01] M. Naor and B. Pinkas. Efficient oblivious transfer protocols. In *Symposium on Discrete Algorithms (SODA'01)*, pages 448–457. ACM/SIAM, 2001.
- [PSSW09] B. Pinkas, T. Schneider, N. P. Smart, and S. C. Williams. Secure two-party computation is practical. In *Advances in Cryptology – ASIACRYPT'09*, volume 5912 of *LNCS*, pages 250–267. Springer, 2009.
- [PVW08] C. Peikert, V. Vaikuntanathan, and B. Waters. A framework for efficient and composable oblivious transfer. In *Advances in Cryptology – CRYPTO'08*, volume 5157 of *LNCS*, pages 554–571. Springer, 2008.
- [Rab81] M. O. Rabin. *How to exchange secrets with oblivious transfer*, TR-81 edition, 1981. Aiken Computation Lab, Harvard University.
- [Sch] P. Scholl. Private communication.
- [SS13] A. Shelat and C.-H. Shen. Fast two-party secure computation with minimal assumptions. In *ACM Computer and Communications Security (CCS'13)*, pages 523–534. ACM, 2013.
- [Yao86] A. C. Yao. How to generate and exchange secrets. In *Foundations of Computer Science (FOCS'86)*, pages 162–167. IEEE, 1986.

A Implementation Details

In this section we provide details about the architecture of our implementation (§A.1), the method we use to allow block-wise evaluation of our protocol and [NNOB12] (§A.2), and discuss the benefits of the pairwise-comparison method described in Protocol 2 (§A.3).

A.1 Architecture

We designed the architecture of the active-secure OT extension implementations such that the communication-intensive passive-secure OT extension routine and the computation-intensive checks on receiver side are performed by separate threads and can be further parallelized independently of each other. This architecture allows us to instantiate the implementation specifically to the available resources of the deployment scenario. More detailed, we can perform the communication-intensive operations with as many threads as required to fully utilize the bandwidth and can then focus the remaining processing power on the computationally-intensive operations. This kind of parallelization offers benefits especially for deployment scenarios of OT extension with small bandwidth, where the network is the bottle-neck for OT extension and where further parallelization of communication-intensive operations would only result in congestion on the network interface. Although this architecture favors our protocol which is computationally more intensive than the protocols of [IKNP03] and [NNOB12], we argue that it nicely fits to today’s increasing number of CPU cores.

A.2 3-Step OT Extension

Note that in order to allow block-wise evaluation of our protocol and [NNOB12], the base-OTs have to be renewed. For the block-wise evaluation of $m \times OT_n$ in b blocks of width w bits ($b = \lceil \frac{m}{w} \rceil$), we perform a 3-step OT extension: In the first step, we perform $\ell \times OT_{b\ell}$ base-OTs using the protocol of [PVW08]. In the second step, we extend $\ell \times OT_{b\ell}$ to $b\ell \times OT_w$ using the respective active secure OT extension protocol. In the third step, we again perform the OT extension step b -times on each ℓ -bit interval, i.e., we extend $\ell \times OT_w$ to $w \times OT_n$ b -times and thereby obtain $bw \geq m$ OTs on n -bit strings.

A.3 Advantages of the Pairwise Comparison Protocol

Although the pairwise comparison Protocol 2 with parameter (168;14,028) is the slowest in our evaluation in §4.2, we stress that it has several advantages which make it favorable in settings with high computation power. The main advantage is that the receiver can pre-compute all checks directly after the base-OTs, since all combinations are checked and hence the sender does not need to send a mapping to the receiver. Additionally, if a computationally powerful device such as a GPU is present, the receiver can use it for computing the checks in parallel.

B Active Secure OT Extension of [NNOB12]

In Protocol 4 we depict the actively-secure OT extension protocol of [NNOB12] with optimizations from [FN13].

PROTOCOL 4 (Active secure OT extension protocol of [NNOB12])

- **Input of P_S :** m pairs (x_j^0, x_j^1) of n -bit strings, $1 \leq j \leq m$.
- **Input of P_R :** m selection bits $\mathbf{r} = (r_1, \dots, r_m)$.
- **Common Input:** Symmetric security parameter κ and $\ell = \lceil \frac{8}{3}\kappa \rceil$.
- **Oracles and primitives:** Ideal $\ell \times OT_\kappa$ functionality, pseudorandom generator G , correlation-robust function H , and random-oracle H' .

1. *Initial OT Phase:*

- (a) P_S initializes a random vector $\mathbf{s} = (s_1, \dots, s_\ell) \in \{0, 1\}^\ell$ and P_R chooses ℓ pairs of seeds $\mathbf{k}_i^0, \mathbf{k}_i^1$ each of size κ .
- (b) The parties invoke the $\ell \times OT_\kappa$ -functionality, where P_S acts as the *receiver* with input \mathbf{s} and P_R acts as the *sender* with inputs $(\mathbf{k}_i^0, \mathbf{k}_i^1)$ for every $1 \leq i \leq \ell$.

For every $1 \leq i \leq \ell$, let $\mathbf{t}^i = G(\mathbf{k}_i^0)$. Let $T = [\mathbf{t}^1 | \dots | \mathbf{t}^\ell]$ denote the $m \times \ell$ bit matrix where its i th column is \mathbf{t}^i for $1 \leq i \leq \ell$. Let \mathbf{t}_j denote the j th row of T for $1 \leq j \leq m$.

2. *OT Extension Phase:*

- (a) P_R computes $\mathbf{t}^i = G(\mathbf{k}_i^0)$ and $\mathbf{u}^i = \mathbf{t}^i \oplus G(\mathbf{k}_i^1) \oplus \mathbf{r}$, and sends \mathbf{u}^i to P_S for every $1 \leq i \leq \ell$.
- (b) For every $1 \leq i \leq \ell$, P_S defines $\mathbf{q}^i = (s_i \cdot \mathbf{u}^i) \oplus G(\mathbf{k}_i^{s_i})$. $\mathbf{q}^i = (s_i \cdot \mathbf{r}) \oplus \mathbf{t}^i$.

3. *Consistency Check of \mathbf{r} :*

- (a) P_S chooses a uniform random permutation $\pi : \{1, \dots, \ell\} \mapsto \{1, \dots, \ell\}$ with $\pi(\pi(i)) = i$ and sends π to Bob. Let $\Pi(\pi) = \{i | i \leq \pi(i)\}$.
- (b) For all $i \in \Pi(\pi)$, P_S computes $d_i = s_i \oplus s_{\pi(i)}$ and $\mathbf{z}^i = \mathbf{q}^i \oplus \mathbf{q}^{\pi(i)}$ sends d_i to P_R .
- (c) P_R computes $\mathbf{z}^i = (d_i \cdot \mathbf{r}) \oplus \mathbf{t}^i \oplus \mathbf{t}^{\pi(i)}$.
- (d) P_S and P_R check equality between $\mathbf{Z} = \mathbf{z}_1 || \dots || \mathbf{z}_{\lfloor \ell/2 \rfloor}$ and $\mathbf{Z}' = \mathbf{z}'_1 || \dots || \mathbf{z}'_{\lfloor \ell/2 \rfloor}$ as follows:
 - i. P_S samples $w \in_R \{0, 1\}^\kappa$, computes $\mathbf{c} = H'(\mathbf{Z} || \mathbf{w})$, sends \mathbf{c} to P_R .
 - ii. P_R then sends \mathbf{Z}' to P_S .
 - iii. P_S checks $\mathbf{Z} \stackrel{?}{=} \mathbf{Z}'$ and aborts on failure. Else sends (\mathbf{Z}, \mathbf{w}) to P_R .
 - iv. P_R checks that $\mathbf{Z} \stackrel{?}{=} \mathbf{Z}'$ and $\mathbf{c} \stackrel{?}{=} H'(\mathbf{Z}' || \mathbf{w})$ and aborts on failure.
- (e) For all $\lfloor \ell/2 \rfloor$ indices in $i \in \Pi(\pi)$ where i is the k th index with $1 \leq k \leq \lfloor \ell/2 \rfloor$, P_S sets $\mathbf{q}'_k = \mathbf{q}_i$ and $s'_k = s_i$ and P_R sets $\mathbf{t}'_k = \mathbf{t}_i$.

4. *OT Extension (continue):*

- (a) Let $Q' = [\mathbf{q}'^1 | \dots | \mathbf{q}'^{\lfloor \ell/2 \rfloor}]$ denote the $m \times \lfloor \ell/2 \rfloor$ bit matrix where its i th column is \mathbf{q}'^i . Let \mathbf{q}'_j denote the j th row of the matrix Q' . (Note that $\mathbf{q}'^i = (s'_i \cdot \mathbf{r}) \oplus \mathbf{t}^i$ and $\mathbf{q}'_j = (r_j \cdot \mathbf{s}') \oplus \mathbf{t}'_j$.)
- (b) P_S sends (y_j^0, y_j^1) for every $1 \leq j \leq m$, where $y_j^0 = x_j^0 \oplus H(j, \mathbf{q}'_j)$ and $y_j^1 = x_j^1 \oplus H(j, \mathbf{q}'_j \oplus \mathbf{s}')$.
- (c) For $1 \leq j \leq m$, P_R computes $x_j = y_j^{r_j} \oplus H(j, \mathbf{t}'_j)$.

5. **Output:** P_R outputs $(x_1^{r_1}, \dots, x_m^{r_m})$; P_S has no output.

C Security in the Presence of a Malicious Sender

We prove that our protocols are secure in the presence of a malicious Sender. We have:

Claim C.1 *Protocol 2 is secure in the presence of a malicious sender, assuming that H is κ -min entropy correlation robust functions.*

Proof: We start with the description of the simulator \mathcal{S} .

The simulator \mathcal{S} .

- Upon receiving auxiliary input z , invoke the adversary \mathcal{A} (controlling the corrupted sender) on z .
- The simulator fixes $\mathbf{r} = 0^m$. It then simulates an honest execution of a receiver with input $\mathbf{r}' = 0^m || \tau$ with the adversary \mathcal{A} for a random τ . In particular, it chooses the random keys $\mathbf{k}_i^0, \mathbf{k}_i^1$. Then, it simulates the base OTs functionality for the adversary, and upon receiving the string $\mathbf{s} = (s_1, \dots, s_\ell)$ it sends it $\{\mathbf{k}_i^{s_i}\}_{i=1}^\ell$. It then sends

$$\mathbf{u}^i = G(\mathbf{k}_i^0) \oplus G(\mathbf{k}_i^1) \oplus (0^m || \tau)$$

for every $i = 1, \dots, \ell$. Finally, for every pair $(\alpha, \beta) \in [\ell]^2$, it computes

$$\begin{aligned} h_{\alpha, \beta}^{s_\alpha, s_\beta} &= H(G(\mathbf{k}_\alpha^{s_\alpha}) \oplus G(\mathbf{k}_\beta^{s_\beta})) , \\ h_{\alpha, \beta}^{\overline{s_\alpha}, \overline{s_\beta}} &= H(G(\mathbf{k}_\alpha^{\overline{s_\alpha}}) \oplus G(\mathbf{k}_\beta^{\overline{s_\beta}})) = H(\mathbf{u}^i \oplus \mathbf{u}^j \oplus G(\mathbf{k}_\alpha^{s_\alpha}) \oplus G(\mathbf{k}_\beta^{s_\beta})) \end{aligned}$$

and sets $h_{\alpha, \beta}^{s_\alpha, s_\beta}$ and $h_{\alpha, \beta}^{\overline{s_\alpha}, \overline{s_\beta}}$ uniformly at random of the appropriate length.

- When the adversary sends during the transfer phase the values (y_j^0, y_j^1) for every $j = 1, \dots, m$, the simulator extracts the inputs (x_j^0, x_j^1) using the function H , the matrix T and the string \mathbf{s} . It then sends these inputs to the trusted parties as the input of the corrupted sender.
- The simulator outputs whatever \mathcal{A} outputs, and halts.

We now show that no distinguisher succeeds to distinguish between the output of the ideal execution and the real execution. Essentially, the key idea is to show that for every possible input \mathbf{r} , no distinguisher can distinguish between an honest execution of $\mathbf{r}' = \mathbf{r} || \tau$ and $\mathbf{r}'' = 0^m || \tau$, and that the values $h_{\alpha, \beta}^{s_\alpha, s_\beta}$ and $h_{\alpha, \beta}^{\overline{s_\alpha}, \overline{s_\beta}}$ are distributed uniformly.

We now show that the joint distribution of the outputs of the parties is indistinguishable in the real world and in the ideal world. Towards this end, consider the following sequence of hybrid games:

- **hyb₀**: This is the real execution (in the base OT hybrid model) with the true input \mathbf{r}' . The output of the execution is the output of the adversary (without loss of generality, its view) and the output of the honest receiver.
- **hyb₁**: In this execution, we have a trusted party for computing the output of the receiver. Specifically, let \mathbf{r}' be the input of the honest receiver, we run an execution of the real protocol with the adversary \mathcal{A} and the receiver with input \mathbf{r}' . When the adversary sends the messages

$\{(y_j^0, y_j^1)\}_{j=1}^m$ in the last step of the protocol, we use the string \mathbf{s} that it has sent to the base OT functionality, and use it to extract the inputs $\{(x_j^0, x_j^1)\}_{j=1}^m$, which are sent to the trusted party as the input of the corrupted sender. The execution then replaces the output of the honest receiver with the values $x_j^{r_j}$.

- **hyb₂**: This is like **hyb₁**, where here we change the definition of base OTs functionality. It first receives \mathbf{s} from the corrupted sender, sends it to the honest receiver, which then sends back to the trusted party uniform keys $\mathbf{k}_i^{s_i}$. The receiver also (locally) chooses keys $\mathbf{k}_i^{\overline{s_i}}$. Moreover, we compute the verification hashes $h_{\alpha,\beta}^{\overline{s_\alpha}, s_\beta}$ as $H(\mathbf{u}^i \oplus \mathbf{u}^j \oplus G(\mathbf{k}_\alpha^{s_\alpha}) \oplus G(\mathbf{k}_\beta^{s_\beta}))$ instead of $H(G(\mathbf{k}_\alpha^{\overline{s_\alpha}}) \oplus G(\mathbf{k}_\beta^{\overline{s_\beta}}))$.
- **hyb₃**: This execution is like **hyb₂**, where every pair $(\alpha, \beta) \in [\ell]^2$ we replace $h_{\alpha,\beta}^{s_\alpha, \overline{s_\beta}}, h_{\alpha,\beta}^{\overline{s_\alpha}, s_\beta}$ with uniformly and independent random values of the appropriate length. Note that now, each string $G(\mathbf{k}_i^{s_i})$ appears only in the transmission of \mathbf{u}^i .
- **hyb₄**: Here, we change the input of the receiver when interacting with the adversary, and use the input 0^m instead of the string \mathbf{r} .

It is easy to see that **hyb₀**, **hyb₁** and **hyb₂** have the same output. The two executions **hyb₂** and **hyb₃** are indistinguishable since H is κ -min entropy correlation robust function. In particular, the distinguisher receives the values $\{\mathbf{k}_i^{s_i}\}_{i=1}^\ell, (\mathbf{u}^1, \dots, \mathbf{u}^\ell)$, as well as \mathbf{r} . Together with H , this uniquely determines the values $\{h_{\alpha,\beta}^{s_\alpha, s_\beta}, h_{\alpha,\beta}^{\overline{s_\alpha}, \overline{s_\beta}}\}_{(\alpha,\beta) \in [\ell]^2}$. Moreover, from each \mathbf{u}^i and $G(\mathbf{k}_i^{s_i})$, it can conclude $G(\mathbf{k}_i^{\overline{s_i}}) \oplus (0^m || \boldsymbol{\tau})$ as $\mathbf{u}^i \oplus G(\mathbf{k}_i^{s_i}) \oplus (\boldsymbol{\tau} || 0^\kappa)$. In fact, for every $(\alpha, \beta) \in [\ell]^2$, it can conclude the value $(G(\mathbf{k}_\alpha^{s_\alpha}) \oplus (0^m || \boldsymbol{\tau})) \oplus G(\mathbf{k}_\beta^{s_\beta})$ and $G(\mathbf{k}_\alpha^{s_\alpha}) \oplus (G(\mathbf{k}_\beta^{\overline{s_\beta}}) \oplus (0^m || \boldsymbol{\tau}))$. On the other hand, for every $(\alpha, \beta) \in [\ell]^2$ we have:

$$\begin{aligned} h_{\alpha,\beta}^{s_\alpha, \overline{s_\beta}} &= H(G(\mathbf{k}_\alpha^{s_\alpha}) \oplus G(\mathbf{k}_\beta^{\overline{s_\beta}})) = H(t_{\alpha,\beta}^1 \oplus (0^m || \boldsymbol{\tau})) \\ h_{\alpha,\beta}^{\overline{s_\alpha}, s_\beta} &= H(G(\mathbf{k}_\alpha^{\overline{s_\alpha}}) \oplus G(\mathbf{k}_\beta^{s_\beta})) = H(t_{\alpha,\beta}^2 \oplus (0^m || \boldsymbol{\tau})) \end{aligned}$$

where

$$t_{\alpha,\beta}^1 = \left(G(\mathbf{k}_\alpha^{s_\alpha}) \oplus (G(\mathbf{k}_\beta^{\overline{s_\beta}}) \oplus (0^m || \boldsymbol{\tau})) \right) \quad \text{and} \quad t_{\alpha,\beta}^2 = \left(G(\mathbf{k}_\alpha^{\overline{s_\alpha}}) \oplus (0^m || \boldsymbol{\tau}) \oplus G(\mathbf{k}_\beta^{s_\beta}) \right),$$

and $t_{\alpha,\beta}^1$ and $t_{\alpha,\beta}^2$ are both known to the distinguisher. Since $\boldsymbol{\tau}$ is chosen uniformly from $\{0, 1\}^\kappa$, assuming that H is a κ -min entropy strongly correlation robust function, the hash values $h_{\alpha,\beta}^{s_\alpha, \overline{s_\beta}}$ and $h_{\alpha,\beta}^{\overline{s_\alpha}, s_\beta}$ are distributed uniformly in the respective domain.

As for **hyb₃** and **hyb₄**, for every i the key $G(\mathbf{k}_i^{\overline{s_i}})$ is a one-time-pad for the string \mathbf{r} , and therefore hides it due to the pseudorandom property of G . ■