On the Optimal Communication Complexity of Error-Correcting Multi-Server PIR

Reo Eriguchi[1,4], Kaoru Kurosawa[2,4], and Koji Nuida[3,4]

[1] Graduate School of Information Science and Technology,
The University of Tokyo, Tokyo, Japan
reo-eriguchi@g.ecc.u-tokyo.ac.jp

[2] Research and Development Initiative, Chuo University, Tokyo, Japan
kaoru.kurosawa.kk@vc.ibaraki.ac.jp

[3] Institute of Mathematics for Industry, Kyushu University, Fukuoka, Japan
nuida@imi.kyushu-u.ac.jp

[4] National Institute of Advanced Industrial Science and Technology, Tokyo, Japan

Abstract. An ℓ-server Private Information Retrieval (PIR) scheme enables a client to retrieve a data item from a database replicated among ℓ servers while hiding the identity of the item. It is called b-error-correcting if a client can correctly compute the data item even in the presence of b malicious servers. It is known that b-error correction is possible if and only if $\ell > 2b$. In this paper, we first prove that if error correction is perfect, i.e., the client always corrects errors, the minimum communication cost of b-error-correcting ℓ-server PIR is asymptotically equal to that of regular $(\ell - 2b)$-server PIR as a function of the database size n. Secondly, we formalize a relaxed notion of statistical b-error-correcting PIR, which allows non-zero failure probability. We show that as a function of n, the minimum communication cost of statistical b-error-correcting ℓ-server PIR is asymptotically equal to that of regular $(\ell - b)$-server one, which is at most that of $(\ell - 2b)$-server one. Our main technical contribution is a generic construction of statistical b-error-correcting ℓ-server PIR for any $\ell > 2b$ from regular $(\ell - b)$-server PIR. We can therefore reduce the problem of determining the optimal communication complexity of error-correcting PIR to determining that of regular PIR. In particular, our construction instantiated with the state-of-the-art PIR schemes and the previous lower bound for single-server PIR result in a separation in terms of communication cost between perfect and statistical error correction for any $\ell > 2b$.

1 Introduction

Private Information Retrieval (PIR) scheme [8] involves a client holding a search index $\tau \in [n]$ and ℓ servers sharing a database $\boldsymbol{a} = (a_1, \dots, a_n) \in \{0,1\}^n$. The scheme enables the client to fetch its desired bit a_τ from the database while hiding the client's index τ from the servers. A trivial solution is to ask a server to send the entire database $\boldsymbol{a}$, which has communication cost $\Theta(n)$. When $\ell = 1$, the trivial solution cannot be improved since it was shown in [8] that a single-server PIR must have communication cost $\Omega(n)$. To achieve communication cost $o(n)$, Chor et al. [8] considered ℓ-server PIR schemes for $\ell \geq 2$ in which servers do not collude. More generally, a PIR scheme is called t-private if any coalition of t servers learns no information on τ. Since then, many ℓ-server PIR schemes have been developed to improve communication cost [1, 3, 4, 7–10, 13, 21].

As more servers are involved, there is a higher possibility that servers are malicious or faulty, or that the databases are not updated simultaneously. It is then important to enable a client to correct errors when part of the servers return false answers. Beimel and Stahl [5] introduced b-error-correcting PIR, in which a client can obtain a correct value a_τ even if b (or less) servers return false answers. Note that they only considered perfect error correction, which requires that a client corrects errors with probability 1. They showed that perfect b-error correction is possible if and only if $\ell > 2b$. In particular, they proposed a generic construction of a perfect b-error-correcting ℓ-server PIR scheme from any $(\ell - 2b)$-server PIR scheme for any $\ell > 2b$.

It has remained an open problem: what is the optimal communication complexity of b-error-correcting ℓ-server PIR as a function of n? For perfect error correction, since the previous construction [5] preserves t-privacy and asymptotic communication cost, the optimal communication complexity of b-error-correcting ℓ-server PIR is asymptotically upper bounded by that of $(\ell - 2b)$-server one. To the best of our knowledge, we have not seen work that studies the optimal communication complexity of *statistical* error-correcting PIR, in which non-zero failure probability is allowed.[5] It is unknown whether we can realize statistical error-correcting PIR with strictly lower communication cost than the perfect one. In this paper, we concern the following problems: (1) Is the minimum communication cost of perfect b-error-correcting ℓ-server PIR asymptotically equal to that of $(\ell - 2b)$-server PIR? (2) What if small failure probability is allowed?

1.1 Our Results

We show answers to the above problems.

1. The optimal communication complexity of perfect b-error-correcting ℓ-server PIR is asymptotically equal to that of $(\ell - 2b)$-server PIR as a function of the database size n.
2. We formalize a relaxed notion of statistical b-error-correcting PIR. The optimal communication complexity of statistical b-error-correcting ℓ-server PIR is asymptotically equal to that of $(\ell - b)$-server PIR as a function of n.

In conclusion, we can reduce the problem of determining the optimal communication complexity of error-correcting PIR to determining that of regular PIR.

[5] For regular PIR, the authors of [12, 18] introduced the statistical analogue of perfect correctness to derive lower bounds for the communication cost of two-server PIR. Statistical correctness allows a client to output an incorrect value with small probability even if all servers behave honestly (Definition 2).

As a corollary, we obtain a separation in terms of communication cost between perfect and statistical error correction. For $\ell > 2b$,

- The minimum communication cost of perfect b-error-correcting $(\ell - 2b)$-private ℓ-server PIR is $\Omega(n)$ since it is equal to that of $(\ell - 2b)$-private $(\ell - 2b)$-server one, which is $\Omega(n)$ [8].[6]
- The minimum communication cost of statistical b-error-correcting $(\ell - 2b)$-private ℓ-server PIR is $o(n)$ since it is equal to that of $(\ell - 2b)$-private $(\ell - b)$-server one, which can be instantiated with the scheme [19].

Optimal Communication Complexity for Perfect Error Correction. We show that perfect b-error-correcting ℓ-server PIR implies regular $(\ell - 2b)$-server PIR with the same communication cost (Theorem 4). Combined with the results of [5], the optimal communication complexity of perfect b-error-correcting ℓ-server PIR is asymptotically equal to that of regular $(\ell - 2b)$-server PIR as a function of n (Corollary 2).

Optimal Communication Complexity for Statistical Error Correction. We show that even statistical b-error-correcting ℓ-server PIR is impossible if $\ell \leq 2b$ (Theorem 5). For $\ell > 2b$, we show that as a function of n, the optimal communication complexity of statistical b-error-correcting ℓ-server PIR is asymptotically equal to that of regular $(\ell - b)$-server PIR with statistical correctness (Corollary 4).

Technically, it follows from our generic transformations preserving t-privacy and asymptotic communication complexity between regular, error-detecting, and error-correcting PIR. Error-detecting PIR [11] is a relaxed notion of error-correcting one in a sense that a client can only detect the existence of errors. We first provide a transformation from statistical regular k-server PIR to statistical b-error-detecting k-server one for any $k > b$ with communication overhead $(\log \epsilon^{-1})^2 k^{4+o(1)}$, where ϵ is the failure probability (Corollary 1). Next, we transform b-error-detecting $(\ell - b)$-server PIR to statistical b-error-correcting ℓ-server one with communication overhead $\binom{\ell}{b}$ (Theorem 3). We therefore obtain a transformation from statistical regular $(\ell - b)$-server PIR to statistical b-error-correcting ℓ-server one with communication overhead $(\log \epsilon^{-1})^2 2^{2\ell + o(\ell)}$ (Corollary 3). Since the overhead is independent of n, our transformation preserves asymptotic communication cost as a function of n. Although it is exponential in ℓ, the overhead is not significant from a practical point of view since the number of servers is typically small, e.g., $\ell = 3$ [7, 10, 13, 21]. Finally, we show that statistical b-error-correcting ℓ-server PIR implies statistical regular $(\ell - b)$-server one with the same communication cost (Theorem 6).

Instantiation of Our Transformation. Since all of the state-of-the-art schemes satisfy perfect correctness, we show a more communication-efficient transfor-

[6] Note that ℓ-private ℓ-server PIR is equivalent to single-server PIR since all the ℓ servers are allowed to collude and hence can be viewed as a single server.

mation that is tailored to perfect regular PIR than Corollary 3 (see Corollary 6). Applying it to [9], we obtain a 1-private statistical b-error-correcting ℓ-server scheme with communication cost $\mathcal{L}_n[r^{-1}, v_k] \cdot \log \epsilon^{-1}$, where $k = \ell - b$, $r = \lfloor \log k \rfloor + 1$ and v_k is a constant depending on k.[7] Based on [19], we obtain a t-private statistical b-error-correcting ℓ-server scheme with communication cost $n^{\lfloor (2k-1)/t \rfloor^{-1}} (\log \epsilon^{-1}) 2^{\ell + o(\ell)}$ for any $t \geq 1$. We also provide a non-generic construction of error-correcting PIR tailored to the ones satisfying a certain algebraic property (Theorem 7). We then obtain a t-private error-correcting scheme with communication cost $n^{\lfloor (2k-1)/t \rfloor^{-1}} (\log n + \log \epsilon^{-1}) \ell^{O(1)}$ for any $t \geq 1$. Note that these t-private schemes are incomparable since the complexity of the latter is polynomial in ℓ while as a function of n, it is larger than the former by a factor of $\log n$.

For any $\ell > 2b$, Corollary 7 gives statistical b-error-correcting $(\ell - 2b)$-private schemes with $o(n)$ communication, while any perfect b-error-correcting $(\ell - 2b)$-private scheme has $\Omega(n)$ communication since we show that it must be based on single-server PIR. This shows a separation in terms of communication cost.

Table 1. Our statistical b-error-correcting t-private ℓ-server PIR schemes for $\ell > 2b$. Let n denote the database size and ϵ denote the failure probability. Let $k = \ell - b$, $r = \lfloor \log k \rfloor + 1$ and v_k denote a constant depending on k.

Method	Communication	t-Privacy	Reference
Corollary 6 + [9]	$\mathcal{L}_n[r^{-1}, v_k] \cdot \log \epsilon^{-1}$	$t = 1$	Corollary 7
Corollary 6 + [19]	$n^{\lfloor (2k-1)/t \rfloor^{-1}} (\log \epsilon^{-1}) 2^{\ell + o(\ell)}$	$t \geq 1$	Corollary 7
Theorem 7 + [19]	$n^{\lfloor (2k-1)/t \rfloor^{-1}} (\log n + \log \epsilon^{-1}) \ell^{O(1)}$	$t \geq 1$	Corollary 8

1.2 Related Work

The scheme [19] is a t-private perfect b-error-correcting PIR scheme with communication cost $n^{\lfloor (2k-1)/t \rfloor^{-1}} \ell^{O(1)}$, where $k = \ell - 2b$. Kurosawa [14] proposed a more time-efficient error correction algorithm for the scheme [19]. On the other hand, the generic construction of [5] instantiated with [19] leads to a perfect error-correcting scheme with communication cost $n^{\lfloor (2k-1)/t \rfloor^{-1}} 2^{O(k)} \ell^{O(1)}$. Although the former has smaller communication cost, they have the same complexity as a function of n.

Error-correcting PIR schemes are considered in the setting where the size of each block of a database is large (see [2, 16, 17, 20] and references therein). Since only the download cost is of interest, the schemes are incomparable with those considered in this paper, where total communication cost is of interest.

[7] We define $\mathcal{L}_n[s, c] = \exp(c (\log n)^s (\log \log n)^{1-s})$ for $0 \leq s \leq 1$ and $c > 0$ (see Section 3).

Eriguchi et al. [11] considered t-private b-error-detecting PIR in a model in which t out of b malicious servers can collude, while we consider a stronger model in which all b malicious servers can collude, which is the same as the one in [5].

The authors of [12, 18] considered regular PIR with statistical correctness and derived lower bounds for the communication cost of two-server PIR. We note that there is no known separation in terms of communication cost between perfect and statistical regular PIR in contrast to our separation for error-correcting PIR.

2 Technical Overview

In this section, we provide an overview of our techniques. We give more detailed descriptions and formal proofs in the following sections.

2.1 Optimal Communication Complexity of Error-Correcting PIR

The Case of Perfect Error Correction. First, we consider perfect error-correcting PIR. Beimel and Stahl [5] showed a generic construction of perfect b-error-correcting ℓ-server PIR from any k-server PIR preserving asymptotic communication complexity, where $k = \ell - 2b$. To determine the optimal communication complexity, we prove the converse of their results: any perfect b-error-correcting ℓ-server PIR scheme Π implies a regular k-server PIR scheme Π' with the same communication cost (Theorem 4).

Let C denote a client with a search index τ and $\mathsf{S}_1, \ldots, \mathsf{S}_\ell$ denote ℓ servers sharing a database $\boldsymbol{a} \in \{0,1\}^n$. For simplicity, we here set $\ell = 3$ and $b = 1$, and assume that S_1 is honest. For a fixed query by C in Π, let $\mathsf{ans}_i(\boldsymbol{a}')$ denote the (deterministic) answer that is generated by S_i when S_i is honest and has database $\boldsymbol{a}'$. We can see that the two sets $\mathcal{X}_h := \{\mathsf{ans}_1(\boldsymbol{a}') \mid \boldsymbol{a}' \text{ satisfies } a'_\tau = h\}$ ($h \in \{0,1\}$) have empty intersection. Then C can determine a_τ solely from a given $\mathsf{ans}_1(\boldsymbol{a})$, which implies single-server PIR, since only the $\mathcal{X}_{a_\tau}$ contains $\mathsf{ans}_1(\boldsymbol{a})$. To see that $\mathcal{X}_0 \cap \mathcal{X}_1 = \emptyset$, assume the contrary, i.e., $\alpha := \mathsf{ans}_1(\boldsymbol{a}') = \mathsf{ans}_1(\boldsymbol{a}'')$ with $a'_\tau = 0$ and $a''_\tau = 1$. Then a malicious server S_2 with database $\boldsymbol{a}'$ can falsely answer $\mathsf{ans}_2(\boldsymbol{a}'')$, yielding the tuple of answers $\vec{\mathsf{ans}} := (\alpha, \mathsf{ans}_2(\boldsymbol{a}''), \mathsf{ans}_3(\boldsymbol{a}'))$. On the other hand, a malicious server S_3 with database $\boldsymbol{a}''$ can falsely answer $\mathsf{ans}_3(\boldsymbol{a}')$, yielding the same tuple of answers $\vec{\mathsf{ans}}$. Now C cannot determine with certainty from $\vec{\mathsf{ans}}$ which of $\boldsymbol{a}'$ or $\boldsymbol{a}''$ was actually used, contradicting the perfect error correction of Π. See Section 7.1 for the details.

The Case of Statistical Error Correction. Next, we consider statistical b-error-correcting PIR. We show equivalence among statistical regular k-server PIR, statistical b-error-detecting k-server PIR, and statistical b-error-correcting ℓ-server PIR, where $k = \ell - b$. We mean by equivalence that a PIR scheme can be transformed to another preserving asymptotic communication complexity as a function of the database size n, and vice versa. Our results on the optimal communication complexity immediately follow from the equivalence between regular k-server PIR and statistical b-error-correcting ℓ-server PIR.

From Regular to Error-Detecting PIR. Our transformation from any statistical k-server PIR scheme $\mathit{\Pi}_0$ to a statistical b-error-detecting k-server PIR scheme $\mathit{\Pi}'$ (Corollary 1) is obtained by composing the following three transformations:

1. From a statistical k-server PIR scheme $\mathit{\Pi}_0$ to a k-server PIR scheme $\mathit{\Pi}_1$ with sufficiently small error probability (Lemma 1). This is done by repeating $\mathit{\Pi}_0$ λ times for some λ and taking the majority of the outputs; now the error probability is negligible in λ due to the Chernoff bound.
2. From $\mathit{\Pi}_1$ to a b-error-detecting k-server PIR scheme $\mathit{\Pi}_2$ where the correctness error probability (i.e., for the case of all honest servers) is sufficiently small and the error detection failure probability is smaller than 1 (see below).
3. From $\mathit{\Pi}_2$ to $\mathit{\Pi}'$ where the error detection failure probability is also negligible (Theorem 2). This is done by repeating $\mathit{\Pi}_2$ λ' times for some λ' and letting the final output be a bit h if all the λ' outputs are h, otherwise $\perp$ meaning "error detected". Now, to fool $\mathit{\Pi}'$, a malicious adversary needs to fool all the λ' instances of $\mathit{\Pi}_2$; due to the structure of $\mathit{\Pi}_2$, it is possible with exponentially small probability in λ' (when ignoring the negligible correctness error probability of $\mathit{\Pi}_2$).

We explain the second transformation from $\mathit{\Pi}_1$ to $\mathit{\Pi}_2$ (Theorem 1), for simplicity with $k = 2$ and $b = 1$. Assume that S_1 is honest. In $\mathit{\Pi}_2$, the client C first randomly guess which of S_1 and S_2 is honest. Suppose that C correctly guesses (with probability 1/2) that S_1 is honest. Secondly, together with a *true* instance of $\mathit{\Pi}_1$, C runs a *dummy* instance of $\mathit{\Pi}_1$ where the query for S_1 is replaced with the same query as S_2. In the dummy instance, an answer returned by a honest server S_1 tells C the correct answer which S_2 should provide if she is honest (we note that each honest server's answer is supposed to be deterministic). Then C runs the two instances in a random order; given servers' answers, C first checks if S_2's answer in the dummy instance is correct (otherwise outputs $\perp$) and then outputs the output in the true instance. Now a malicious server S_2 who wants to fool $\mathit{\Pi}_2$ has to correctly guess which is the dummy instance, honestly behave in the dummy instance, and modify the answer in the true instance. Since the two instances are executed in a random order and are indistinguishable from S_2's viewpoint, S_2 can guess correctly with probability at most 1/2. In summary, C can detect error with probability at least $(1/2) \cdot (1/2) = 1/4$ (when ignoring the negligible correctness error probability of $\mathit{\Pi}_1$), while the correctness error probability of $\mathit{\Pi}_2$ is almost the same as that of $\mathit{\Pi}_1$ since $\mathit{\Pi}_2$ runs only two instances of $\mathit{\Pi}_1$. See Section 5.1 for the details of the above method.

By carefully adjusting the parameters λ and λ' in the above transformations, we can make the error probability of the final scheme $\mathit{\Pi}'$ bounded by a given value $\epsilon_{\mathrm{ED}} > 0$. If the communication cost of the initial scheme $\mathit{\Pi}_0$ is c_0, that of $\mathit{\Pi}'$ is $c = c_0(\log \epsilon_{\mathrm{ED}}^{-1})^2 \cdot \mathsf{poly}(k)$, which is asymptotically equal to c_0 as a function of n. See Section 5.2 for the details.

From Error-Detecting to Error-Correcting PIR. Our transformation from any statistical b-error-detecting k-server PIR scheme $\mathit{\Pi}'$ to a statistical b-error-correcting ℓ-server PIR scheme $\mathit{\Pi}$ simply executes $N := \binom{\ell}{\ell-b}$ independent instances of $\mathit{\Pi}'$, each interacting with one of the N subsets of $k = \ell - b$ servers (Theorem 3). The output of Π is any bit contained in the N outputs by Π' if it exists; otherwise $\perp$. Now when all the N instances of Π' work correctly, the N outputs contain at least one true bit a_τ in the instance with k honest servers, due to the correctness of Π'; and do not contain the opposite bit due to the error-detection capability of the other $N - 1$ instances. Therefore Π fails only if some of the N instances of Π' fails, which happens with probability at most N times larger than the failure probability of Π'. Note that we can make the failure probability of Π arbitrarily small by starting from Π' with sufficiently small failure probability. See Section 6 for the details.

From Error-Correcting to Regular PIR. Finally, we prove that any statistical b-error-correcting ℓ-server PIR scheme Π implies a statistical regular k-server PIR scheme with the same error probability and communication complexity (Theorem 6). This is simply done as follows: When the client C receives correct answers in Π from $k = \ell - b$ honest servers only, C feds b arbitrary answers to the reconstruction algorithm of Π. Since those b answers can be viewed as false answers by the remaining b malicious servers, C correctly retrieves an item due to the error correction capability of Π. See Section 7.2 for the details.

2.2 Instantiation of Our Transformation

We instantiate our transformation from regular to statistical error-correcting PIR with the state-of-the-art schemes [9, 19]. Although the above transformation (Corollary 3) can be used, we show a construction tailored to regular PIR with perfect correctness since the schemes [9, 19] are perfectly correct. We observe that if a regular scheme is perfect, we do not need to make correctness error negligible at the first step of our construction of error-detecting PIR. We show a more communication-efficient construction than the above one (Corollary 6). Instantiated with [9, 19], it gives the statistical error-correcting schemes shown in the first and second rows of Table 1. Note that if a statistical regular PIR scheme advances state of the art in the future, we should use the transformation in Corollary 3 instead of that in Corollary 6.

The third scheme can be obtained as follows. The construction follows the framework of the Rabin-BenOr robust secret sharing scheme [15]. Their scheme uses tags produced by a message authentication code (MAC, for short) to verify the integrity of shares. Since a client needs to verify the authenticity of *computations* by servers in PIR, we use a homomorphic MAC of [6]. The answer computed by any honest server is accepted by all honest servers while any incorrect answer is detected by them. If $\ell > 2b$, at least $\ell - b$ answers are declared to be correct after verification procedures. A client runs the reconstruction algorithm of $(\ell - b)$-server PIR on them and obtains a correct value. Note that the tag size of [6] grows linearly in the depth of the evaluated arithmetic circuit. The above construction is non-generic in a sense that it requires that server-side computation is represented by a shallow arithmetic circuit. Since the scheme [19] satisfies it, we obtain the third scheme of Table 1. See Section 8.1 for the details.

3 Preliminaries

For $m \in \mathbb{N}$, define $[m] = \{1, \ldots, m\}$. For a subset X of a set Y, we define $\overline{X} = \{y \in Y : y \notin X\}$ if Y is clear from the context. We write $u \leftarrow_{\$} Y$ if u is chosen uniformly at random from a set Y. Define $\binom{[m]}{k}$ as the set of all subsets of $[m]$ of size k. Let $\mathfrak{R}_{\mathcal{A}}$ denote the set of all random strings for a probabilistic algorithm $\mathcal{A}$. Namely, on input x, $\mathcal{A}$ outputs $\mathcal{A}(x; r)$ for $r \leftarrow_{\$} \mathfrak{R}_{\mathcal{A}}$. For a vector $\boldsymbol{x}$, let x_i denote the i-th entry of $\boldsymbol{x}$. Let $\log x$ denote the base-2 logarithm of x and $\ln x$ denote the base-e logarithm of x, where e denotes the Napier's constant. Let $\mathcal{L}_n[s, c]$ denote the function of n defined as $\mathcal{L}_n[s, c] = \exp(c(\log n)^s(\log\log n)^{1-s})$, where $0 \leq s \leq 1$ and $c > 0$.

4 Private Information Retrieval (PIR)

4.1 Definitions

Definition 1 (Syntax). *An ℓ-server PIR scheme Π for a universe of databases $\{0,1\}^n$ consists of three algorithms $\Pi = (\mathcal{Q}, \mathcal{A}, \mathcal{D})$, where $\mathcal{Q}$ is probabilistic while $\mathcal{A}$ and $\mathcal{D}$ are deterministic:*

- *A query algorithm $\mathcal{Q}$ takes a search index $\tau \in [n]$ as input. It then samples a random string $r \leftarrow_{\$} \mathfrak{R}_{\mathcal{Q}}$ and outputs $\mathsf{que}_i \in \{0,1\}^{c_{\mathsf{que}}}$ for $i \in [\ell]$ and $\mathsf{aux} \in \{0,1\}^{c_{\mathsf{aux}}}$. That is, $\mathcal{Q}(\tau; r) = (\mathsf{que}_1, \ldots, \mathsf{que}_\ell; \mathsf{aux})$.*
- *An answer algorithm $\mathcal{A}$ takes $i \in [\ell]$, $\mathsf{que}_i \in \{0,1\}^{c_{\mathsf{que}}}$ and $\boldsymbol{a} = (a_1, \ldots, a_n) \in \{0,1\}^n$ as input and outputs $\mathsf{ans}_i \in \{0,1\}^{c_{\mathsf{ans}}}$. That is, $\mathcal{A}(i, \mathsf{que}_i, \boldsymbol{a}) = \mathsf{ans}_i$.*
- *A reconstruction algorithm $\mathcal{D}$ takes $(\mathsf{ans}_1, \ldots, \mathsf{ans}_\ell) \in (\{0,1\}^{c_{\mathsf{ans}}})^\ell$ and $\mathsf{aux} \in \{0,1\}^{c_{\mathsf{aux}}}$ as input, and outputs $y \in \{0,1\}$. That is, $\mathcal{D}(\mathsf{ans}_1, \ldots, \mathsf{ans}_\ell; \mathsf{aux}) = y$.*

The (total) communication complexity of Π is given by $\ell(c_{\mathsf{que}} + c_{\mathsf{ans}})$.

Definition 2 (Statistical correctness). *An ℓ-server PIR scheme $\Pi = (\mathcal{Q}, \mathcal{A}, \mathcal{D})$ is said to be $(1-\epsilon)$-correct if for any $\boldsymbol{a} = (a_1, \ldots, a_n) \in \{0,1\}^n$ and any $\tau \in [n]$, it holds that $\Pr[r \leftarrow_{\$} \mathfrak{R}_{\mathcal{Q}} : \mathcal{D}(\mathsf{ans}_1, \ldots, \mathsf{ans}_\ell; \mathsf{aux}) = a_\tau] \geq 1 - \epsilon$, where $(\mathsf{que}_1, \ldots, \mathsf{que}_\ell; \mathsf{aux}) = \mathcal{Q}(\tau; r)$ and $\mathsf{ans}_i = \mathcal{A}(i, \mathsf{que}_i, \boldsymbol{a})$ for $i \in [\ell]$.*

Remark 1. In the literature, a PIR scheme is usually required to satisfy perfect correctness, i.e., $\epsilon = 0$. We use the above generalized notion of $(1-\epsilon)$-correctness for $\epsilon \geq 0$ to show the equivalence between $(1-\epsilon)$-correct $(\ell - b)$-server PIR and $(b; 1-\epsilon)$-error-correcting ℓ-server PIR.

Definition 3 (t-Privacy). *An ℓ-server PIR scheme $\Pi = (\mathcal{Q}, \mathcal{A}, \mathcal{D})$ is said to be t-private if for any $X \in \binom{[\ell]}{t}$ and any $\tau, \tau' \in [n]$, the distributions of $(\mathsf{que}_i)_{i \in X}$ and $(\mathsf{que}'_i)_{i \in X}$ are perfectly identical, where $r, r' \leftarrow_{\$} \mathfrak{R}_{\mathcal{Q}}$, $(\mathsf{que}_1, \ldots, \mathsf{que}_\ell; \mathsf{aux}) = \mathcal{Q}(\tau; r)$ and $(\mathsf{que}'_1, \ldots, \mathsf{que}'_\ell; \mathsf{aux}') = \mathcal{Q}(\tau'; r')$.*

4.2 Robust PIR

(k, ℓ)-Robust PIR [5] guarantees that a client can compute a_τ from answers of any k out of ℓ servers. We provide a general notion of $(k, \ell; 1-\epsilon)$-robust PIR with statistical correctness.

Definition 4. *A PIR scheme $\Pi = (\mathcal{Q}, \mathcal{A}, \mathcal{D})$ is said to be $(k, \ell; 1-\epsilon)$-robust if*

- *$\mathcal{D}$ takes $X \in \binom{[\ell]}{k}$, $(\mathsf{ans}_i)_{i \in X} \in (\{0,1\}^{c_{\mathsf{ans}}})^k$ and $\mathsf{aux} \in \{0,1\}^{c_{\mathsf{aux}}}$ as input, and outputs $y \in \{0,1\}$;*
- *It holds that $\Pr[r \leftarrow_\$ \mathfrak{R}_\mathcal{Q} : \mathcal{D}(X, (\mathsf{ans}_i)_{i \in X}; \mathsf{aux}) = a_\tau] \geq 1-\epsilon$ for any $\boldsymbol{a} \in \{0,1\}^n$, any $\tau \in [n]$ and any $X \in \binom{[\ell]}{k}$, where $(\mathsf{que}_1, \ldots, \mathsf{que}_\ell; \mathsf{aux}) = \mathcal{Q}(\tau; r)$ and $\mathsf{ans}_i = \mathcal{A}(i, \mathsf{que}_i, \boldsymbol{a})$ for $i \in [\ell]$.*

4.3 Error-Correcting and Error-Detecting PIR

We can identify an ℓ-server PIR scheme Π with a protocol $(\Pi; \mathsf{C}, \mathsf{S}_1, \ldots, \mathsf{S}_\ell)$ between a client C and ℓ servers $\mathsf{S}_1, \ldots, \mathsf{S}_\ell$ as follows:

Query. On input $\tau \in [n]$, C chooses $r \leftarrow_\$ \mathfrak{R}_\mathcal{Q}$ and computes $(\mathsf{que}_1, \ldots, \mathsf{que}_\ell; \mathsf{aux}) = \mathcal{Q}(\tau; r)$. Then, C sends que_i to S_i for $i \in [\ell]$.
Answer. On input $\boldsymbol{a} \in \{0,1\}^n$, each S_i returns $\mathsf{ans}_i = \mathcal{A}(i, \mathsf{que}_i, \boldsymbol{a})$ to C.
Reconstruction. C outputs $y = \mathcal{D}(\mathsf{ans}_1, \ldots, \mathsf{ans}_\ell; \mathsf{aux})$.

We consider a malicious adversary $\mathcal{B}$ who corrupts a set B of at most b servers and returns a possibly modified answer $\widetilde{\mathsf{ans}}_i$ to C instead of ans_i for each $i \in B$.[8]

Definition 5 (Error-correcting PIR). *A PIR scheme Π is said to be $(b; 1-\epsilon_{\mathrm{EC}})$-error-correcting if for any $\boldsymbol{a} \in \{0,1\}^n$, any $\tau \in [n]$ and any malicious adversary $\mathcal{B}$ who corrupts at most b servers, it holds that $\Pr[\mathsf{C}$ outputs $a_\tau] \geq 1-\epsilon_{\mathrm{EC}}$ in the protocol $(\Pi; \mathsf{C}, \mathsf{S}_1, \ldots, \mathsf{S}_\ell)$.*

Remark 2. In the definition, in order to achieve stronger error correction capability, we allow the modified answers to depend on all of the b queries even if $b > t$ for t-privacy (though now b queries may leak some information on the client's index). This model follows the original definition in [5].

Definition 6 (Error-detecting PIR). *A PIR scheme Π is said to be $(b; 1-\epsilon_{\mathrm{ED}})$-error-detecting if the following conditions hold:*

- *Π is $(1-\epsilon_{\mathrm{ED}})$-correct.*
- *C is allowed to output a special symbol $\perp$ and for any $\boldsymbol{a} \in \{0,1\}^n$, any $\tau \in [n]$ and any malicious adversary $\mathcal{B}$ who corrupts at most b servers, it holds that $\Pr[\mathsf{C}$ outputs $a_\tau] \geq 1-\epsilon_{\mathrm{ED}}$ in the protocol $(\Pi; \mathsf{C}, \mathsf{S}_1, \ldots, \mathsf{S}_\ell)$.*

[8] More formally, we formalize $\mathcal{B}$ by using a *tampering function* [11]. See [11] or Appendix A for the details.

5 Transformation from Regular to Error-Detecting PIR

We show a generic transformation from any t-private k-server PIR scheme $\mathit{\Pi}_0 = (\mathcal{Q}_0, \mathcal{A}_0, \mathcal{D}_0)$ to a t-private $(k-1; 1-\epsilon_{\mathrm{ED}})$-error-detecting k-server PIR scheme $\mathit{\Pi}$. The communication overhead is independent of the database size n.

We first give our transformation for larger ϵ_{ED} in Section 5.1, which is then reduced to arbitrarily small $\epsilon_{\mathrm{ED}} > 0$ in Section 5.2.

5.1 Basic Transformation

Given $\mathit{\Pi}_0 = (\mathcal{Q}_0, \mathcal{A}_0, \mathcal{D}_0)$, we consider the following two query algorithms $\mathit{\Pi}_0^{\mathsf{Compute}}$ and $\mathit{\Pi}_0^{\mathsf{Verify},(i,j)}$. $\mathit{\Pi}_0^{\mathsf{Compute}}$ is used to actually compute a_τ and $\mathit{\Pi}_0^{\mathsf{Verify},(i,j)}$ is used to verify whether S_j correctly computes her answer assuming that S_i is honest.

$\mathit{\Pi}_0^{\mathsf{Compute}}$: On input $\tau \in [n]$, C chooses $r \leftarrow_\$ \mathfrak{R}_{\mathcal{Q}_0}$ and computes $\mathcal{Q}_0(\tau; r) = (\mathsf{que}_1, \ldots, \mathsf{que}_k; \mathsf{aux})$. Then, he sends (m, que_m) to S_m for $m \in [k]$.

$\mathit{\Pi}_0^{\mathsf{Verify},(i,j)}$: On input $\tau \in [n]$, C chooses $r \leftarrow_\$ \mathfrak{R}_{\mathcal{Q}_0}$ and computes $\mathcal{Q}_0(\tau; r) = (\mathsf{que}_1, \ldots, \mathsf{que}_k; \mathsf{aux})$. Then, he sends (m, que_m) to S_m for $m \in [k] \setminus \{i\}$, and (j, que_j) to S_i.

Now we consider a k-server PIR scheme $\mathit{\Pi}_1 = (\mathcal{Q}, \mathcal{A}, \mathcal{D})$ as shown in Fig. 1, where the client C chooses $i \neq j \in [k]$ uniformly at random, randomly permutes two instances of $\mathit{\Pi}_0^{\mathsf{Compute}}$ and $\mathit{\Pi}_0^{\mathsf{Verify},(i,j)}$, and executes them in parallel with $\mathsf{S}_1, \ldots, \mathsf{S}_k$. C verifies that S_j correctly computes her answer using $\mathit{\Pi}_0^{\mathsf{Verify},(i,j)}$ and then he runs $\mathcal{D}_0$ on the answers obtained during the execution of $\mathit{\Pi}_0^{\mathsf{Compute}}$.

We obtain the following theorem. We sketch the proof here. The formal proof is given in Appendix B, where a more general result (Theorem 2) is proved.

Theorem 1. *If $\mathit{\Pi}_0$ is t-private and $(1-\epsilon)$-correct, then $\mathit{\Pi}_1$ is t-private and $(k-1; 1-\epsilon_{\mathrm{ED}})$-error-detecting for $\epsilon_{\mathrm{ED}} = 1 - 1/(2k(k-1)) + 2\epsilon$. Furthermore, if $\mathit{\Pi}_0$ has communication cost c_0, then $\mathit{\Pi}_1$ has communication cost $O(c_0 + k \log k)$.*

Proof (Sketch). It is easy to see that $\mathit{\Pi}_1$ is $(1-\epsilon)$-correct. It is also easy to see that $\mathit{\Pi}_1$ is t-private and has the communication complexity $2c_0 + O(k \log k) = O(c_0 + k \log k)$. We will prove that $\mathit{\Pi}_1$ is $(k-1; 1-\epsilon_{\mathrm{ED}})$-error-detecting.

First, we assume that $\epsilon = 0$. Without loss of generality, we suppose that S_1 is honest and a malicious adversary $\mathcal{B}$ corrupts $\mathsf{S}_2, \ldots, \mathsf{S}_k$. Clearly, if $\mathsf{S}_2, \ldots, \mathsf{S}_k$ return correct answers, the client C obtains the correct value a_τ. We may assume that at least one malicious server, say S_2, modifies her answer.

Consider the case where C chooses $(i, j) = (1, 2)$ at Step 1(a) in Fig. 1, which occurs with probability $1/(k(k-1))$. To make C output the incorrect value $1 - a_\tau$, S_2 needs to honestly behave in the instance $\mathit{\Pi}_0^{\mathsf{Verify},(1,2)}$ and to modify her answer in the other instance $\mathit{\Pi}_0^{\mathsf{Compute}}$. Note that $\mathcal{B}$ cannot distinguish between two instances $\mathit{\Pi}_0^{\mathsf{Compute}}, \mathit{\Pi}_0^{\mathsf{Verify},(1,2)}$ since the distributions of queries that

Components.

- A k-server PIR scheme $\mathit{\Pi}_0 = (\mathcal{Q}_0, \mathcal{A}_0, \mathcal{D}_0)$
- Query algorithms $\mathit{\Pi}_0^{\mathsf{Compute}}$ and $\mathit{\Pi}_0^{\mathsf{Verify},(i,j)}$

Query. On input $\tau \in [n]$, C chooses $i \neq j \in [k]$ uniformly at random and executes $\mathit{\Pi}_0^{\mathsf{Compute}}, \mathit{\Pi}_0^{\mathsf{Verify},(i,j)}$ in a random order. Specifically, C does the following:

1. He chooses $i \neq j \in [k]$ uniformly at random.
2. He randomly permutes two protocols $\mathit{\Pi}_0^{\mathsf{Compute}}, \mathit{\Pi}_0^{\mathsf{Verify},(i,j)}$. Let $\mathit{\Pi}_0^{(1)}, \mathit{\Pi}_0^{(2)}$ denote the resulting sequence.
3. He generates queries for $\mathit{\Pi}_0^{(1)}, \mathit{\Pi}_0^{(2)}$. Let $\mathsf{que}_m^{(\alpha)}$ and $\mathsf{aux}^{(\alpha)}$ denote the query sent to S_m and auxiliary information obtained during the execution of $\mathit{\Pi}_0^{(\alpha)}$, respectively.
4. He sends $\mathsf{que}_m = (\mathsf{que}_m^{(1)}, \mathsf{que}_m^{(2)})$ to each S_m.

Answer. On input $\boldsymbol{a} \in \{0,1\}^n$, each S_m does the following:

1. For each $\mathsf{que}_m^{(\alpha)} = (x, \mathsf{que}_x)$, she computes $\mathsf{ans}_m^{(\alpha)} = \mathcal{A}_0(x, \mathsf{que}_x, \boldsymbol{a})$.
2. She returns $\mathsf{ans}_m^{(\alpha)}$ to C for $\alpha \in \{1,2\}$.

Error detection. If he receives $\widetilde{\mathsf{ans}}_m^{(\alpha)}$ from S_m as $\mathsf{ans}_m^{(\alpha)}$, C does the following:

1. For $\alpha \in \{1,2\}$ with $\mathit{\Pi}_0^{(\alpha)} = \mathit{\Pi}_0^{\mathsf{Compute}}$, he sets $z \leftarrow \mathcal{D}_0(\widetilde{\mathsf{ans}}_1^{(\alpha)}, \ldots, \widetilde{\mathsf{ans}}_k^{(\alpha)}; \mathsf{aux}^{(\alpha)})$.
2. For $\alpha \in \{1,2\}$ with $\mathit{\Pi}_0^{(\alpha)} = \mathit{\Pi}_0^{\mathsf{Verify},(i,j)}$, he verifies whether $\widetilde{\mathsf{ans}}_j^{(\alpha)} = \widetilde{\mathsf{ans}}_i^{(\alpha)}$ holds. If it holds, then he outputs z. Otherwise, he outputs $\perp$.

Fig. 1. A basic error-detecting PIR protocol $\mathit{\Pi}_1$

$\mathsf{S}_2, \ldots, \mathsf{S}_k$ receive are the same in both cases. Hence, the distribution of an answer returned by S_2 is independent of the permutation chosen by C at Step 1(b). With probability at least 1/2, S_2 fails to guess the instance $\mathit{\Pi}_0^{\mathsf{Verify},(1,2)}$, in which she has to behave honestly. Therefore, C can detect errors with probability $1/(2k(k-1)) = 1 - \epsilon_{\mathrm{ED}}$. We conclude that $\mathit{\Pi}_1$ is $(k-1; 1-\epsilon_{\mathrm{ED}})$-error-detecting.

In the general case of $\epsilon \geq 0$, the previous argument still holds unless C chooses a *bad* random string such that $\mathcal{D}_0$ outputs $1 - a_\tau$ even if all servers return correct answers. Since $\mathit{\Pi}_1$ involves two instances of $\mathit{\Pi}_0$ and $\mathit{\Pi}_0$ is $(1-\epsilon)$-correct, we can upper bound by 2ϵ the fraction of such bad random strings. Therefore the previous bound for the error probability is increased by 2ϵ, which results in the value of ϵ_{ED} in the statement. □

5.2 General Transformation

We consider a PIR scheme $\mathit{\Pi}$ obtained by running sufficiently many independent instances of the basic error-detecting scheme (Fig. 2). We obtain the following theorem. The formal proof is given in Appendix B.

Theorem 2. *Let $b < k$, $\lambda \in \mathbb{N}$, $\epsilon \geq 0$, and $\epsilon_{\mathrm{ED}} = 2\lambda\epsilon + (1 - 1/(2k(k-1)))^\lambda$. If there exists a $(1-\epsilon)$-correct t-private k-server PIR scheme $\mathit{\Pi}_0 = (\mathcal{Q}_0, \mathcal{A}_0, \mathcal{D}_0)$ with communication cost c_0, then there exists a $(b; 1-\epsilon_{\mathrm{ED}})$-error-detecting t-private k-server PIR scheme with communication cost $c = O(\lambda(c_0 + k \log k))$.*

1. For each $\nu \in [\lambda]$, run the protocol Π_1 (Fig. 1) in parallel.
2. Let $z^{(\nu)}$ denote the output of C for $\nu \in [\lambda]$.
3. If $\{z^{(\nu)} : \nu \in [\lambda]\} = \{z\}$ for some $z \in \{0,1\}$, C outputs z. Otherwise C outputs $\perp$.

Fig. 2. A general error-detecting PIR Protocol Π

Proof (Sketch). Since Π runs λ instances of Π_1, the claim for the communication complexity follows from Theorem 1. When $\epsilon = 0$, a malicious adversary has to fool all the λ independent instances in order to make the client output an incorrect value. Therefore the claimed bound follows from that of Theorem 1. In the general case of $\epsilon \geq 0$, since Π runs Π_0 2λ times in total, the bound is increased by $2\lambda\epsilon$. □

We show that the optimal communication complexity of $(k-1; 1-\epsilon_{ED})$-error-detecting k-server PIR is asymptotically upper bounded by that of $(1-\epsilon_0)$-correct k-server PIR for any $\epsilon_{\mathrm{ED}} > 0$ and $\epsilon_0 < 1/2$. Lemma 1 shows that the correctness error of a $(1-\epsilon_0)$-correct PIR scheme can be made arbitrarily small. We defer the formal proof to Appendix C.

Lemma 1. *Let $0 \leq \epsilon_0 < 1/2$ and $\lambda \in \mathbb{N}$. If there exists a $(1-\epsilon_0)$-correct t-private k-server PIR scheme Π_0 with communication cost c_0, then there exists a $(1-\epsilon)$-correct t-private k-server PIR scheme Π with communication cost $c_0\lambda$, where $\epsilon = (2\sqrt{\epsilon_0(1-\epsilon_0)})^\lambda \leq \exp(-2(1/2-\epsilon_0)^2\lambda)$.*

Proof (Sketch). We construct Π in a way that it runs λ independent instances of Π_0 and takes the majority of their outputs. Since each instance fails with probability at most ϵ_0, the Chernoff bound implies that the majority fails with probability at most ϵ as in the statement. The claimed upper bound for ϵ is deduced by an elementary analysis. □

Corollary 1 shows a general transformation from any $(1-\epsilon_0)$-correct k-server PIR scheme to a $(k-1; 1-\epsilon_{\mathrm{ED}})$-error-detecting PIR scheme.

Corollary 1. *Let $b < k$, $0 \leq \epsilon_0 < 1/2$ and $\epsilon_{\mathrm{ED}} > 0$. If there exists a $(1-\epsilon_0)$-correct t-private k-server PIR scheme with communication cost c_0, then there exists a $(b; 1-\epsilon_{\mathrm{ED}})$-error-detecting t-private k-server PIR scheme with communication cost $c = c_0(\log \epsilon_{\mathrm{ED}}^{-1})^2(1/2-\epsilon_0)^{-2}k^{4+o(1)}$.*

Proof. Let $\lambda, \lambda' \in \mathbb{N}$ be the smallest integers such that $\lambda \geq 2k(k-1)(\log 3\epsilon_{\mathrm{ED}}^{-1})$ and $\lambda' \geq (1/2-\epsilon_0)^{-2}\lambda$. We have that $\lambda = O(k^2 \log \epsilon_{\mathrm{ED}}^{-1})$ and $\lambda' = O((1/2-\epsilon_0)^{-2}\lambda)$. Let $\epsilon = (2\sqrt{\epsilon_0(1-\epsilon_0)})^{\lambda'} \leq \exp(-2(1/2-\epsilon_0)^2\lambda') \leq \exp(-2\lambda)$. It follows from Lemma 1 that there exists a $(1-\epsilon)$-correct t-private k-server PIR scheme with communication cost $c_0' = c_0\lambda'$. Note that

$$\begin{aligned} 2\lambda\epsilon + \left(1 - \frac{1}{2k(k-1)}\right)^{\lambda} &\leq 2\exp(-\lambda) + \left(1 - \frac{1}{2k(k-1)}\right)^{\lambda} \\ &\leq 3\exp\left(-\frac{\lambda}{2k(k-1)}\right) \\ &\leq \epsilon_{\mathrm{ED}}. \end{aligned}$$

It then follows from Theorem 2 that there exists a $(b; 1-\epsilon_{\mathrm{ED}})$-error-detecting t-private k-server PIR scheme with communication cost $c = O(\lambda(c_0' + k\log k)) = c_0(\log \epsilon_{\mathrm{ED}}^{-1})^2(1/2 - \epsilon_0)^{-2}k^{4+o(1)}$. □

6 Transformation from Error-Detecting to Error-Correcting PIR

We show a transformation from any $(b; 1-\epsilon_{\mathrm{ED}})$-error-detecting $(\ell-b)$-server PIR scheme to a $(b; 1-\epsilon_{\mathrm{EC}})$-error-correcting ℓ-server PIR scheme. Our transformation simply executes $\binom{\ell}{\ell-b}$ independent instances of the error-detecting PIR scheme. In particular, the communication overhead is independent of the database size n. We sketch the proof here but refer to Appendix D for the details.

Theorem 3. *Let $b < \ell/2$, $k = \ell - b$ and $N = \binom{\ell}{k}$. If there exists a $(b; 1-\epsilon_{\mathrm{ED}})$-error-detecting t-private k-server PIR scheme Π_0 with communication cost c , then there exists a $(b; 1-\epsilon_{\mathrm{EC}})$-error-correcting t-private ℓ-server PIR scheme with communication cost Nc for $\epsilon_{\mathrm{EC}} = N\epsilon_{\mathrm{ED}}$.*

Proof (Sketch). We consider a PIR scheme Π where N independent instances of Π_0 are executed between a client and every subset of k servers. Let $z_1, \ldots, z_N \in \{0, 1, \perp\}$ be the N outcomes. If $\{z_1, \ldots, z_N\}$ is $\{s\}$ or $\{s, \perp\}$ for some $s \in \{0, 1\}$, then the client outputs s and otherwise outputs 0.

Clearly, the communication complexity of Π is Nc. It is also easy to see that Π is t-private since all executions of Π_0 are independent.

For the correctness, a malicious adversary can make the output incorrect only if either the unique instance of Π_0 with k honest servers does not output a_τ (happening with probability at most ϵ_{ED}) or some of the other $N-1$ instances of Π_0 with possibly corrupted servers fails to detect error (happening with probability at most ϵ_{ED} each). Therefore the error probability is bounded by $\epsilon_{\mathrm{ED}} + (N-1)\epsilon_{\mathrm{ED}} = \epsilon_{\mathrm{EC}}$. □

7 Optimal Communication Complexity of Error-Correcting PIR

In this section, we show the relation between the optimum communication complexity of error-correcting PIR and that of regular PIR as a function of the database size n. We use the following notations: For the database size n,

- $\mathrm{PIR}_{k,t;1-\epsilon}(n)$ denotes the minimum communication cost of t-private $(1-\epsilon)$-correct k-server PIR schemes and;
- EC-$\mathrm{PIR}_{\ell,t,b;1-\epsilon}(n)$ denotes the minimum communication cost of t-private $(b;1-\epsilon)$-error-correcting ℓ-server PIR schemes.

7.1 The Case of Perfect Error Correction

Beimel and Stahl [5] showed a generic transformation from a t-private 1-correct k-server PIR scheme to a t-private 1-correct $(b;1)$-error-correcting ℓ-server PIR scheme for $b \leq (\ell-k)/2$. The communication overhead is $2^{O(k)}\ell\log\ell$, which is independent of the database size n. We show the converse.

Theorem 4. *Let $b < \ell/2$. If there exists a $(b;1)$-error-correcting ℓ-server PIR scheme $\Pi = (\mathcal{Q},\mathcal{A},\mathcal{D})$, then there exists a deterministic algorithm $\mathcal{D}'$ such that $\Pi' = (\mathcal{Q},\mathcal{A},\mathcal{D}')$ is a $(k,\ell;1)$-robust PIR scheme, where $k=\ell-2b$.*

Proof. Let $X \in \binom{[\ell]}{k}$ and $\tau \in [n]$. Below we show that for any possible output $((\mathsf{que}_i)_{i\in[\ell]};\mathsf{aux})$ of $\mathcal{Q}(\tau)$, if two databases $\boldsymbol{a}'$ and $\boldsymbol{a}''$ satisfy $a'_\tau = 0$ and $a''_\tau = 1$, then we always have $(\mathcal{A}(i,\mathsf{que}_i,\boldsymbol{a}'))_{i\in X} \neq (\mathcal{A}(i,\mathsf{que}_i,\boldsymbol{a}''))_{i\in X}$. Once this is proved, the desired $\mathcal{D}'$ can be constructed as follows: Given $(\mathsf{ans}_i)_{i\in X}$, it first finds (by an exhaustive search) a database $\widehat{\boldsymbol{a}} \in \{0,1\}^n$ such that $(\mathcal{A}(i,\mathsf{que}_i,\widehat{\boldsymbol{a}}))_{i\in X} = (\mathsf{ans}_i)_{i\in X}$, and then outputs $z = \widehat{a}_\tau$. Note that the actual database $\widehat{\boldsymbol{a}} = \boldsymbol{a}$ indeed satisfies the equality and the uniqueness yields $z = a_\tau$.

Now we show the claim. Assume for the contrary that $(\mathcal{A}(i,\mathsf{que}_i,\boldsymbol{a}'))_{i\in X} = (\mathcal{A}(i,\mathsf{que}_i,\boldsymbol{a}''))_{i\in X}$, which we denote by $(\alpha_i)_{i\in X}$. We fix a partition $[\ell] = X \cup Y \cup Z$ into mutually disjoint parts with $Y, Z \in \binom{[\ell]}{b}$. For each $i \in [\ell]$, define $\widetilde{\alpha}_i \in \{0,1\}^{c_{\mathsf{ans}}}$ by $\widetilde{\alpha}_i = \alpha_i$ if $i \in X$, $\widetilde{\alpha}_i = \mathcal{A}(i,\mathsf{que}_i,\boldsymbol{a}')$ if $i \in Y$, and $\widetilde{\alpha}_i = \mathcal{A}(i,\mathsf{que}_i,\boldsymbol{a}'')$ if $i \in Z$. Now if Y (resp. Z) is the set of corrupted servers, then a malicious adversary with database $\boldsymbol{a}''$ (resp. $\boldsymbol{a}'$) can let the tuple of answers be $\widetilde{\alpha}$ by setting $\widetilde{\mathsf{ans}}_i = \mathcal{A}(i,\mathsf{que}_i,\boldsymbol{a}')$ for $i \in Y$ (resp. $\mathcal{A}(i,\mathsf{que}_i,\boldsymbol{a}'')$ for $i \in Z$). Therefore, the perfect correctness of Π implies that $\mathcal{D}(\widetilde{\alpha};\mathsf{aux})$ must output $a''_\tau = 1$ (resp. $a'_\tau = 0$). This is a contradiction. Therefore the claim holds. □

A t-private $(k,\ell;1)$-robust PIR scheme trivially implies a t-private 1-correct k-server PIR scheme. By combining the results of [5] and Theorem 4, we obtain the following corollary.

Corollary 2. *For any $b < \ell/2$ and $t \geq 1$, it holds that*

$$\text{EC-PIR}_{\ell,t,b;1}(n) = \Theta_{\ell,b}\left(\mathrm{PIR}_{\ell-2b,t;1}(n)\right),$$

where the notation $\Theta_{\ell,b}(\cdot)$ hides any constant depending on ℓ and b.

7.2 The Case of Statistical Error Correction

In [5], it is claimed that $(b;1)$-error-correcting ℓ-server PIR is impossible if $b \geq \ell/2$. We show a more general impossibility result in the case of statistical correctness. The proof is given in Appendix E.

Theorem 5. *Let $b \geq \ell/2$. If there exists a $(b; 1-\epsilon_{\mathrm{EC}})$-error-correcting ℓ-server PIR scheme $\Pi = (\mathcal{Q}, \mathcal{A}, \mathcal{D})$, then $\epsilon_{\mathrm{EC}} \geq 1/2$.*

For $b < \ell/2$, we obtain a generic construction of $(b; 1-\epsilon)$-error-correcting ℓ-server PIR from $(1-\epsilon)$-correct $(\ell-b)$-server PIR by combining Corollary 1 and Theorem 3.

Corollary 3. *Let $b < \ell/2$, $k = \ell - b$, $0 < \epsilon_0 < 1/2$ and $\epsilon > 0$. If there exists a $(1-\epsilon_0)$-correct t-private k-server PIR scheme with communication cost c_0, then there exists a $(b; 1-\epsilon)$-error-correcting t-private ℓ-server PIR scheme with communication cost $c = c_0(\log \epsilon^{-1})^2(1/2-\epsilon_0)^{-2}2^{\ell+o(\ell)}$.*

Proof. Let $\epsilon_{\mathrm{ED}} = \epsilon/\binom{\ell}{b}$. Corollary 1 implies that there exists a $(b; 1-\epsilon_{\mathrm{ED}})$-error-detecting t-private k-server PIR scheme with communication cost $c_1 = c_0(\log \epsilon_{\mathrm{ED}}^{-1})^2(1/2-\epsilon_0)^{-2}k^{4+o(1)} = c_0(\log \epsilon^{-1})^2(1/2-\epsilon_0)^{-2}\ell^{O(1)}$. Then, Theorem 3 implies that there exists a $(b; 1-\epsilon)$-error-correcting t-private ℓ-server PIR scheme with communication cost $c = c_1\binom{\ell}{b} = c_0(\log \epsilon^{-1})^2(1/2-\epsilon_0)^{-2}2^{\ell+o(\ell)}$. □

The following theorem shows the converse of Corollary 3.

Theorem 6. *Let $b < \ell/2$ and $\epsilon_{\mathrm{EC}} \geq 0$. If there exists a $(b; 1-\epsilon_{\mathrm{EC}})$-error-correcting ℓ-server PIR scheme $\Pi = (\mathcal{Q}, \mathcal{A}, \mathcal{D})$, there exists an algorithm $\mathcal{D}'$ such that $\Pi' = (\mathcal{Q}, \mathcal{A}, \mathcal{D}')$ is a $(k, \ell; 1-\epsilon_{\mathrm{EC}})$-robust PIR scheme, where $k = \ell - b$.*

Proof. Let $X \in \binom{[\ell]}{k}$, $\tau \in [n]$ and $\boldsymbol{a} \in \{0,1\}^n$. Given X, $(\widetilde{\mathsf{ans}}_i)_{i \in X}$, and aux as input, $\mathcal{D}'$ is defined in a way that it sets $\widetilde{\mathsf{ans}}_i = \mathbf{0} \in \{0,1\}^{c_{\mathsf{ans}}}$ for $i \in \overline{X}$ and runs $\mathcal{D}((\widetilde{\mathsf{ans}}_i)_{i \in [\ell]}; \mathsf{aux})$. Now the input for the internal $\mathcal{D}$ is equivalent to the case of Π where the database is $\boldsymbol{a}$ and a malicious adversary corrupting the servers in $\overline{X}$ has modified each $\widetilde{\mathsf{ans}}_i$, $i \in \overline{X}$ to $\mathbf{0}$. Therefore the correctness of $\mathcal{D}$ implies that $\mathcal{D}'$ outputs a_τ with probability at least $1-\epsilon_{\mathrm{EC}}$. □

A t-private $(k, \ell; 1-\epsilon)$-robust PIR scheme trivially implies a t-private $(1-\epsilon)$-correct k-server PIR scheme. By combining Corollary 3 and Theorem 6, we obtain the following corollary.

Corollary 4. *For any $b < \ell/2$, $t \geq 1$ and $0 < \epsilon < 1/2$, it holds that*

$$\text{EC-PIR}_{\ell,t,b;1-\epsilon}(n) = \Theta_{\ell,b,\epsilon}\left(\text{PIR}_{\ell-b,t;1-\epsilon}(n)\right),$$

where the notation $\Theta_{\ell,b,\epsilon}(\cdot)$ hides any constant depending on ℓ, b and ϵ.

8 Instantiation of Our Transformation

We have shown the generic construction of statistical error-correcting PIR from regular PIR with *statistical* correctness. The reason is that we aim at relating its optimal communication complexity to that of statistical regular PIR. Since all the state-of-the-art schemes satisfy *perfect* correctness, we use a construction tailored to perfect regular PIR in the following instantiations. If an initial scheme is 1-correct, the resulting error-correcting scheme has better communication cost than Corollary 3.

The following corollary shows that if an initial scheme is 1-correct, it is possible to construct a more efficient error-detecting PIR scheme than Corollary 1.

Corollary 5. *Let $b < k$ and $\epsilon_{\mathrm{ED}} > 0$. If there exists a 1-correct t-private k-server PIR scheme with communication cost c_0, then there exists a $(b; 1 - \epsilon_{\mathrm{ED}})$-error-detecting t-private k-server PIR scheme with communication cost $c = O((c_0 + k \log k)k^2(\log \epsilon_{\mathrm{ED}}^{-1}))$.*

Proof. Let $\lambda \in \mathbb{N}$ be the smallest integer such that $\lambda \geq 2k(k-1)\log \epsilon_{\mathrm{ED}}^{-1}$. Observe that $\lambda = O(k^2 \log \epsilon_{\mathrm{ED}}^{-1})$. Also, observe that $(1 - 1/(2k(k-1)))^{\lambda} \leq \exp(-\lambda/(2k(k-1))) \leq \epsilon_{\mathrm{ED}}$. The statements then follow from Theorem 2. □

By combining Theorem 3 with Corollary 5, we obtain the following corollary.

Corollary 6. *Let $b < \ell/2$, $k = \ell - b$ and $\epsilon > 0$. If there exists a 1-correct t-private k-server PIR scheme with communication cost c_0, then there exists a $(b; 1-\epsilon)$-error-correcting t-private ℓ-server PIR scheme with communication cost $c = c_0(\log \epsilon^{-1})2^{\ell + o(\ell)}$.*

Proof. Let $\epsilon_{\mathrm{ED}} = \epsilon / \binom{\ell}{b}$. Corollary 5 implies that there exists a $(b; 1 - \epsilon_{\mathrm{ED}})$-error-detecting t-private k-server PIR scheme with communication complexity $c_1 = c_0(\log \epsilon_{\mathrm{ED}}^{-1})k^{3+o(1)} = c_0(\log \epsilon^{-1})\ell^{O(1)}$. Then, Theorem 3 implies that there exists a $(b; 1 - \epsilon)$-error-correcting t-private ℓ-server PIR scheme with communication complexity $c = c_1\binom{\ell}{b} = c_0(\log \epsilon^{-1})2^{\ell + o(\ell)}$. □

We apply Corollary 6 to the state-of-the-art PIR schemes [9, 19] to obtain the following corollary.

Corollary 7. *Let $b < \ell/2$, $k = \ell - b$, $t \geq 1$ and $\epsilon > 0$. There exist $(b; 1 - \epsilon)$-error-correcting ℓ-server PIR schemes Π_1, Π_2 such that:*

- *Π_1 is 1-private and has communication cost $\mathcal{L}_n[(\lfloor \log k \rfloor + 1)^{-1}, v_k] \cdot (\log \epsilon^{-1})2^{\ell + o(\ell)}$, where v_k is a constant depending only on k and;*
- *Π_2 is t-private and has communication cost $n^{\lfloor (2k-1)/t \rfloor^{-1}}(\log \epsilon^{-1})2^{\ell + o(\ell)}$.*

Observe that if $\ell > b + t$, i.e., $k \geq t + 1 \geq 2$, then Corollary 7 gives t-private $(b; 1 - \epsilon)$-error-correcting ℓ-server PIR with $o(n)$ communication. On the other hand, if $\ell \leq 2b + t$, any $(b; 1)$-error-correcting t-private ℓ-server scheme has communication cost $\Omega(n)$ since it must be based on single-server PIR in view of Corollary 2. Thus, there is a separation in terms of communication cost between perfect and statistical b-error-correcting t-private ℓ-server PIR if $\max\{2b, b + t\} < \ell \leq 2b + t$.

In the next section, we show a non-generic construction assuming a certain algebraic property. If we apply it to the scheme of [19], we obtain a $(b; 1 - \epsilon)$-error-correcting ℓ-server PIR scheme which is more communication-efficient in terms of ℓ than Π_2 in Corollary 7.

8.1 Statistical Error-Correcting PIR Based on Homomorphic MAC

We show a construction of a $(b; 1-\epsilon)$-error-correcting ℓ-server PIR scheme Π from a $(\ell - b, \ell; 1-\epsilon')$-robust PIR scheme $\Pi_0 = (\mathcal{Q}_0, \mathcal{A}_0, \mathcal{D}_0)$. Our construction requires that $\mathcal{A}_0$ is represented by low-degree polynomials. The communication overhead is polynomial in ℓ while that of Corollary 3 is exponential in ℓ.

Our construction is based on the framework of the Rabin-BenOr robust secret sharing scheme [15]. Suppose that the client C wants a server S_i to compute a function $F_{\boldsymbol{a},i}(\cdot) := \mathcal{A}_0(i, \cdot, \boldsymbol{a})$ on input que_i. Since C does not know $\boldsymbol{a}$, he verifies the computation of $F_{\boldsymbol{a},i}(\mathsf{que}_i)$ with help of the other servers S_j $(j \neq i)$. Below we give a specific method to do so based on homomorphic MAC. Our method guarantees that if S_i computes $F_{\boldsymbol{a},i}$ correctly, her answer is accepted by any honest server S_j. Since the number of honest servers $\ell - b$ is greater than that of dishonest ones b, C then knows that S_i is honest. If S_i does not perform computation correctly, she will be detected with high probability by honest servers. Again, since there are more honest servers than dishonest ones, C can discard the answer returned by S_i. After the above procedures, at least $\ell - b$ answers are declared to be correct. The $(\ell - b, \ell; 1-\epsilon)$-robustness implies that C can run $\mathcal{D}_0$ based on those answers.

To verify computation of $F_{\boldsymbol{a},i}$, we uses some techniques for information-theoretic MACs [6]. For simplicity, we assume that $F_{\boldsymbol{a},i}$ is a single polynomial of degree d over a finite field $\mathbb{F}_p$ and a query is a field element $\mathsf{que}_i \in \mathbb{F}_p$. The client C chooses a random field element α_{ij} from a sufficiently large field $\mathbb{F}_q$, which is used as a secret key to verify the computation of S_i with help of S_j. The authentication tag of the message que_i is a random polynomial $T_{ij}(X)$ of degree 1 over $\mathbb{F}_q$ that evaluates to que_i on the point 0. C sends que_i and T_{ij} to S_i, and $\rho_{ij} = T_{ij}(\alpha_{ij})$ to S_j. Then, S_i computes $\mathsf{ans}_i = F_{\boldsymbol{a},i}(\mathsf{que}_i)$ and also $G_{ij}(X) = F_{\boldsymbol{a},i}(T_{ij}(X))$ while S_j computes $\sigma_{ij} = F_{\boldsymbol{a},i}(\rho_{ij})$. Finally, C verifies whether it holds that $\mathsf{ans}_i = G_{ij}(0)$ and $\sigma_{ij} = G_{ij}(\alpha_{ij})$. Even if S_i sends an incorrect answer $\widetilde{\mathsf{ans}}_i$ along with a modified tag $\widetilde{G}_{ij}(X)$ such that $\widetilde{\mathsf{ans}}_i = \widetilde{G}_{ij}(0)$, C can detect errors unless α_{ij} happens to be a root of a non-zero polynomial $\widetilde{G}_{ij} - G_{ij}$, which occurs with probability roughly $O(d/q)$. The above argument can be generalized into the case where que_i is a vector over $\mathbb{F}_p$ and $F_{\boldsymbol{a},i}$ is a tuple of multiple polynomials.

Note that the size of each tag grows linearly with the degree of the evaluated polynomial. Since the above verification procedure is performed over every pair of servers, the communication complexity of the resulting scheme Π is $\ell^2 d$ times larger than Π_0 ignoring logarithmic factors of d, ℓ and q. We obtain the following theorem. See Appendix F for the proof.

Theorem 7. *Let $b < \ell/2$, $k = \ell - b$, $\epsilon_0 \geq 0$ and $\epsilon > \binom{\ell}{b}\epsilon_0$. Let $\Pi_0 = (\mathcal{Q}_0, \mathcal{A}_0, \mathcal{D}_0)$ be a $(k, \ell; 1-\epsilon_0)$-robust t-private ℓ-server PIR scheme such that:*

- *For any $i \in [\ell]$, a query que_i is an M-dimensional vector over a finite field $\mathbb{F}_p$, i.e., $\mathsf{que}_i \in \mathbb{F}_p^M$;*

- *For any $\boldsymbol{a} \in \{0,1\}^n$ and any $i \in [\ell]$, $\mathcal{A}_0(i, \cdot, \boldsymbol{a})$ is a tuple $(F^{(\mu)}_{\boldsymbol{a},i})_{\mu \in [N]}$ of M-variate polynomials of total degree at most d over $\mathbb{F}_p$, i.e., $\mathcal{A}_0(i, \mathsf{que}_i, \boldsymbol{a}) = (F^{(\mu)}_{\boldsymbol{a},i}(\mathsf{que}_i))_{\mu \in [N]}$ for $\mathsf{que}_i \in \mathbb{F}_p^M$.*

Then there exists a $(b; 1-\epsilon)$-error-correcting t-private ℓ-server PIR scheme with communication complexity

$$c = O\left(\ell^2(M + Nd) \log \frac{Nd\ell}{\epsilon - \binom{\ell}{b}\epsilon_0}\right).$$

The scheme in [19] satisfies the assumptions of Theorem 7 with $\epsilon_0 = 0$, $p = O(\log \ell)$, $d = \lfloor (2k-1)/t \rfloor$ and $M, N \in O(dn^{1/d})$. Thus we obtain the following corollary.

Corollary 8. *Let $b < \ell/2$, $k = \ell - b$, $t \geq 1$ and $\epsilon > 0$. There exists a t-private $(b; 1-\epsilon)$-error-correcting ℓ-server PIR scheme with communication complexity $n^{\lfloor (2k-1)/t \rfloor^{-1}}(\log n + \log \epsilon^{-1})k^{1+o(1)}\ell^{2+o(1)}$.*

Acknowledgements

This research was partially supported by JSPS KAKENHI Grant Numbers JP20J20797 and 19H01109, Japan, JST CREST Grant Number JPMJCR2113, Japan, and JST AIP Acceleration Research JPMJCR22U5, Japan.

References

1. Ambainis, A.: Upper bound on the communication complexity of private information retrieval. In: Automata, Languages and Programming. pp. 401–407 (1997)
2. Banawan, K., Ulukus, S.: The capacity of private information retrieval from byzantine and colluding databases. IEEE Transactions on Information Theory **65**(2), 1206–1219 (2019)
3. Beimel, A., Ishai, Y., Kushilevitz, E., Raymond, J.F.: Breaking the o(n/sup 1/(2k-1)/) barrier for information-theoretic private information retrieval. In: The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings. pp. 261–270 (2002)
4. Beimel, A., Ishai, Y.: Information-theoretic private information retrieval: A unified construction. In: Automata, Languages and Programming. pp. 912–926 (2001)
5. Beimel, A., Stahl, Y.: Robust information-theoretic private information retrieval. Journal of Cryptology **20**(3), 295–321 (2007)
6. Catalano, D., Fiore, D.: Practical homomorphic MACs for arithmetic circuits. In: Advances in Cryptology – EUROCRYPT 2013. pp. 336–352 (2013)
7. Chee, Y.M., Feng, T., Ling, S., Wang, H., Zhang, L.F.: Query-efficient locally decodable codes of subexponential length. computational complexity **22**(1), 159–189 (2013)
8. Chor, B., Goldreich, O., Kushilevitz, E., Sudan, M.: Private information retrieval. Journal of the ACM **45**(6), 965–982 (1998)

9. Dvir, Z., Gopi, S.: 2-server pir with subpolynomial communication. Journal of the ACM **63**(4), 1–15 (2016)
10. Efremenko, K.: 3-query locally decodable codes of subexponential length. SIAM Journal on Computing **41**(6), 1694–1703 (2012)
11. Eriguchi, R., Kurosawa, K., Nuida, K.: Multi-server PIR with full error detection and limited error correction. In: 3rd Conference on Information-Theoretic Cryptography (ITC 2022). pp. 1:1–1:20 (2022)
12. Goldreich, O., Karloff, H., Schulman, L., Trevisan, L.: Lower bounds for linear locally decodable codes and private information retrieval. computational complexity **15**(3), 263–296 (2006)
13. Itoh, T., Suzuki, Y.: Improved constructions for query-efficient locally decodable codes of subexponential length. IEICE Transactions on Information and Systems **E93.D**(2), 263–270 (2010)
14. Kurosawa, K.: How to correct errors in multi-server PIR. In: Advances in Cryptology – ASIACRYPT 2019. pp. 564–574 (2019)
15. Rabin, T., Ben-Or, M.: Verifiable secret sharing and multiparty protocols with honest majority. In: Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing. pp. 73–85. STOC '89 (1989)
16. Sun, H., Jafar, S.A.: The capacity of private information retrieval. IEEE Transactions on Information Theory **63**(7), 4075–4088 (2017)
17. Sun, H., Jafar, S.A.: The capacity of robust private information retrieval with colluding databases. IEEE Transactions on Information Theory **64**(4), 2361–2370 (2018)
18. Wehner, S., de Wolf, R.: Improved lower bounds for locally decodable codes and private information retrieval. In: Automata, Languages and Programming. pp. 1424–1436 (2005)
19. Woodruff, D., Yekhanin, S.: A geometric approach to information-theoretic private information retrieval. In: 20th Annual IEEE Conference on Computational Complexity (CCC'05). pp. 275–284 (2005)
20. Yang, E., Xu, J., Bennett, K.: Private information retrieval in the presence of malicious failures. In: Proceedings 26th Annual International Computer Software and Applications. pp. 805–810 (2002)
21. Yekhanin, S.: Towards 3-query locally decodable codes of subexponential length. Journal of the ACM (JACM) **55**(1), 1–16 (2008)

A Definitions

Following [11], we use the notion of tampering functions to formalize a malicious server who corrupts a set of servers and modifies their answers.

Definition 7 (Tampering function). *Let $\Pi = (\mathcal{Q}, \mathcal{A}, \mathcal{D})$ be an ℓ-server PIR scheme. Let $T \subseteq [\ell]$ be a subset. Let f be a function which takes $(\mathsf{que}_1, \ldots, \mathsf{que}_\ell) \in (\{0,1\}^{c_{\mathsf{que}}})^\ell$ and $\boldsymbol{a} \in \{0,1\}^n$ as input, and outputs $(\widetilde{\mathsf{ans}}_1, \ldots, \widetilde{\mathsf{ans}}_\ell) \in (\{0,1\}^{c_{\mathsf{ans}}})^\ell$. We say that f is a tampering function for Π with respect to T if for each $i \in [\ell]$, it holds that*

$$\widetilde{\mathsf{ans}}_i = \begin{cases} \mathcal{A}(i, \mathsf{que}_i, \boldsymbol{a}), & \text{if } i \notin T, \\ f_i(\{\mathsf{que}_{i'}\}_{i' \in T}, \boldsymbol{a}), & \text{if } i \in T, \end{cases}$$

for some function f_i. We denote the family of all such tampering functions by $\mathcal{F}_T^{\Pi}$.

Definition 8 (Error-correcting PIR). *We say that an ℓ-server PIR scheme $\Pi = (\mathcal{Q}, \mathcal{A}, \mathcal{D})$ is $(1 - \epsilon_{\mathrm{EC}})$-error-correcting with respect to T if for any $\boldsymbol{a} = (a_1, \ldots, a_n) \in \{0,1\}^n$, any $\tau \in [n]$ and any $f \in \mathcal{F}_T^{\Pi}$, it holds that*

$$\Pr[r \leftarrow_{\$} \mathfrak{R}_{\mathcal{Q}} : \mathcal{D}(f(\mathsf{que}_1, \ldots, \mathsf{que}_\ell, \boldsymbol{a}); \mathsf{aux}) = a_\tau] \geq 1 - \epsilon_{\mathrm{EC}},$$

where $(\mathsf{que}_1, \ldots, \mathsf{que}_\ell; \mathsf{aux}) = \mathcal{Q}(\tau; r)$. We say that an ℓ-server PIR scheme Π is $(b; 1 - \epsilon_{\mathrm{EC}})$-error-correcting if it is $(1 - \epsilon_{\mathrm{EC}})$-error-correcting with respect to any $T \subseteq [\ell]$ of size b.

Definition 9 (Error-detecting PIR). *We say that an ℓ-server PIR scheme $\Pi = (\mathcal{Q}, \mathcal{A}, \mathcal{D})$ is $(1 - \epsilon_{\mathrm{ED}})$-error-detecting with respect to T if the following conditions hold:*

- *Π is $(1 - \epsilon_{\mathrm{ED}})$-correct.*
- *$\mathcal{D}$ is allowed to output a special symbol $\perp$ and it holds that for any $\boldsymbol{a} = (a_1, \ldots, a_n) \in \{0,1\}^n$, any $\tau \in [n]$ and any $f \in \mathcal{F}_T^{\Pi}$,*

  $$\Pr[r \leftarrow_{\$} \mathfrak{R}_{\mathcal{Q}} : \mathcal{D}(f(\mathsf{que}_1, \ldots, \mathsf{que}_\ell, \boldsymbol{a}); \mathsf{aux}) \in \{a_\tau, \perp\}] \geq 1 - \epsilon_{\mathrm{ED}},$$

  where $(\mathsf{que}_1, \ldots, \mathsf{que}_\ell; \mathsf{aux}) = \mathcal{Q}(\tau; r)$.

We say that an ℓ-server PIR scheme Π is $(b; 1 - \epsilon_{\mathrm{ED}})$-error-detecting if it is $(1 - \epsilon_{\mathrm{ED}})$-error-detecting with respect to any subset T of size b.

B Proof of Theorem 2

Let $\mathcal{I} = \{(i,j) \in [k]^2 : i \neq j\}$. Let Π be a k-server PIR scheme $\Pi = (\mathcal{Q}, \mathcal{A}, \mathcal{D})$ described in Figs. 3, 4 and 5.

Communication complexity. The communication complexity of Π is at most

$$\lambda(2c_0 + O(\log k)) = O(\lambda(c_0 + \log k)).$$

Correctness. Assume that all servers are honest. Let $\boldsymbol{a} \in \{0,1\}^n$ be a database and $\tau \in [n]$ be a client's index. Let $\nu \in [\lambda]$. We show that the value $z^{(\nu)}$ computed at Step 2(b) of $\mathcal{D}$ is 0 or 1 with probability 1 and is equal to a_τ with probability at least $1-\epsilon$. If so, the union bound implies that it holds that $\{z^{(\nu)} : \nu \in [\lambda]\} = \{a_\tau\}$ with probability at least $1 - \lambda\epsilon$, which shows the $(1 - \epsilon_{\mathrm{ED}})$-correctness of Π.

Assume that $b^{(\nu)} = 0$. We can deal with the other case of $b^{(\nu)} = 1$ similarly. Observe that the first row of $\boldsymbol{Q}^{(\nu)}$ is

$$\mathsf{row}_1^{(j)} = \left((1, \mathsf{que}_{1,1}^{(\nu)}), \ldots, (k, \mathsf{que}_{1,k}^{(\nu)})\right).$$

Since all servers are honest, the first row of $\boldsymbol{A}^{(\nu)}$ is

$$(\widetilde{\mathsf{ans}}_{1,1}^{(\nu)}, \ldots, \widetilde{\mathsf{ans}}_{1,k}^{(\nu)}; \mathsf{aux}_1^{(\nu)}) = \left(\mathcal{A}(1, \mathsf{que}_{1,1}^{(\nu)}, \boldsymbol{a}), \ldots, \mathcal{A}(k, \mathsf{que}_{1,k}^{(\nu)}, \boldsymbol{a}); \mathsf{aux}_1^{(\nu)}\right).$$

```
Notations.
    – A $(1-\epsilon)$-correct $k$-server PIR scheme $\Pi_0 = (\mathcal{Q}_0, \mathcal{A}_0, \mathcal{D}_0)$
    – $\mathcal{I} = \{(i,j) \in [k]^2 : i \neq j\}$

$\mathcal{Q}(\tau)$. Given $\tau \in [n]$:
    1. For each $\nu \in [\lambda]$, do the following:
        (a) Choose $r^{(\nu)} = (r_m^{(\nu)})_{m \in \{1,2\}} \leftarrow_{\$} (\mathfrak{R}_{\mathcal{Q}_0})^2$.
        (b) Choose $(i,j) = (i^{(\nu)}, j^{(\nu)}) \leftarrow_{\$} \mathcal{I}$ and $b^{(\nu)} \leftarrow_{\$} \{0,1\}$.
        (c) Do the following:
            – If $b^{(\nu)} = 0$, set
                $$\mathsf{row}_1^{(\nu)} = \left((1, \mathsf{que}_{1,1}^{(\nu)}), \ldots, (k, \mathsf{que}_{1,k}^{(\nu)})\right),$$
                $$\mathsf{row}_2^{(\nu)} = \left((1, \mathsf{que}_{2,1}^{(\nu)}), \ldots, (j-1, \mathsf{que}_{2,j-1}^{(\nu)}), (i, \mathsf{que}_{2,i}^{(\nu)}), (j+1, \mathsf{que}_{2,j+1}^{(\nu)}), \ldots, (k, \mathsf{que}_{2,k}^{(\nu)})\right),$$
              where $(\mathsf{que}_{m,1}^{(\nu)}, \ldots, \mathsf{que}_{m,k}^{(\nu)}; \mathsf{aux}_m^{(\nu)}) = \mathcal{Q}_0(\tau; r_m^{(\nu)})$.
            – If $b^{(\nu)} = 1$, set
                $$\mathsf{row}_1^{(\nu)} = \left((1, \mathsf{que}_{1,1}^{(\nu)}), \ldots, (j-1, \mathsf{que}_{1,j-1}^{(\nu)}), (i, \mathsf{que}_{1,i}^{(\nu)}), (j+1, \mathsf{que}_{1,j+1}^{(\nu)}), \ldots, (k, \mathsf{que}_{1,k}^{(\nu)})\right),$$
                $$\mathsf{row}_2^{(\nu)} = \left((1, \mathsf{que}_{2,1}^{(\nu)}), \ldots, (k, \mathsf{que}_{2,k}^{(\nu)})\right),$$
              where $(\mathsf{que}_{m,1}^{(\nu)}, \ldots, \mathsf{que}_{m,k}^{(\nu)}; \mathsf{aux}_m^{(\nu)}) = \mathcal{Q}_0(\tau; r_m^{(\nu)})$.
        (d) Construct an 2-by-$k$ matrix $\boldsymbol{Q}^{(\nu)}$ as
                $$\boldsymbol{Q}^{(\nu)} = \begin{pmatrix} \mathsf{row}_1^{(\nu)} \\ \mathsf{row}_2^{(\nu)} \end{pmatrix}.$$
        (e) Let $\mathsf{que}_i^{(\nu)}$ be the $i$-th column of $\boldsymbol{Q}^{(\nu)}$ for $i \in [k]$.
        (f) Let $\mathsf{aux}^{(\nu)} = (\mathsf{aux}_1^{(\nu)}, \mathsf{aux}_2^{(\nu)})$.
    2. Let $\mathsf{que}_i = (\mathsf{que}_i^{(\nu)})_{\nu \in [\lambda]}$ for $i \in [k]$.
    3. Let $\mathsf{aux} = ((\mathsf{aux}^{(\nu)})_{\nu \in [\lambda]}, (i^{(\nu)}, j^{(\nu)}, b^{(\nu)})_{\nu \in [\lambda]})$.
    4. Output $(\mathsf{que}_1, \ldots, \mathsf{que}_k; \mathsf{aux})$.
```

Fig. 3. The query algorithm of the PIR scheme Π in Theorem 2

```
$\mathcal{A}(i, \mathsf{que}_i, \boldsymbol{a})$. Given $i \in [k]$, $\mathsf{que}_i = (\mathsf{que}_i^{(\nu)})_{\nu \in [\lambda]}$ and $\boldsymbol{a} \in \{0,1\}^n$:
  1. For each $\nu \in [\lambda]$, do the following:
     (a) For each $m = 1, 2$, if the $m$-th entry of $\mathsf{que}_i^{(\nu)}$ is $(x_m, \mathsf{que}_{m,x_m}^{(\nu)})$, let $\mathsf{ans}_{m,i}^{(\nu)} =$
         $\mathcal{A}_0(x_m, \mathsf{que}_{m,x_m}^{(\nu)}, \boldsymbol{a})$.
     (b) Let
         $$\mathsf{ans}_i^{(\nu)} = \begin{pmatrix} \mathsf{ans}_{1,i}^{(\nu)} \\ \mathsf{ans}_{2,i}^{(\nu)} \end{pmatrix}.$$
  2. Output $\mathsf{ans}_i = (\mathsf{ans}_i^{(\nu)})_{\nu \in [\lambda]}$.
```

Fig. 4. The answer algorithm of Π

```
$\mathcal{D}(\widetilde{\mathsf{ans}}_1, \ldots, \widetilde{\mathsf{ans}}_k; \mathsf{aux})$. Given $\widetilde{\mathsf{ans}}_i = (\widetilde{\mathsf{ans}}_i^{(\nu)})_{\nu \in [\lambda]}$ $(i \in [k])$, where $\widetilde{\mathsf{ans}}_i^{(\nu)}$ has the same
    form as $\mathsf{ans}_i^{(\nu)}$, and $\mathsf{aux} = ((\mathsf{aux}^{(\nu)})_{\nu \in [\lambda]}, (i^{(\nu)}, j^{(\nu)}, b^{(\nu)})_{\nu \in [\lambda]})$:
  1. $\mathcal{L} = \emptyset$.
  2. For each $\nu \in [\lambda]$, do the following:
     (a) Construct an 2-by-$(k+1)$ matrix $\boldsymbol{A}^{(\nu)}$ as
         $$\boldsymbol{A}^{(\nu)} = \begin{pmatrix} \widetilde{\mathsf{ans}}_1^{(\nu)} & \cdots & \widetilde{\mathsf{ans}}_k^{(\nu)} & \mathsf{aux}^{(\nu)} \end{pmatrix} = \begin{pmatrix} \widetilde{\mathsf{ans}}_{1,1}^{(\nu)} & \cdots & \widetilde{\mathsf{ans}}_{1,k}^{(\nu)} & \mathsf{aux}_1^{(\nu)} \\ \widetilde{\mathsf{ans}}_{2,1}^{(\nu)} & \cdots & \widetilde{\mathsf{ans}}_{2,k}^{(\nu)} & \mathsf{aux}_2^{(\nu)} \end{pmatrix}.$$
     (b) Do the following:
         – If $b^{(\nu)} = 0$:
           i. Compute
              $$y^{(\nu)} = \mathcal{D}_0\left(\widetilde{\mathsf{ans}}_{1,1}^{(\nu)}, \ldots, \widetilde{\mathsf{ans}}_{1,k}^{(\nu)}; \mathsf{aux}_1^{(\nu)}\right).$$
           ii. Check whether $\widetilde{\mathsf{ans}}_{2,i}^{(\nu)} = \widetilde{\mathsf{ans}}_{2,j}^{(\nu)}$ holds. If the equality holds, set $z^{(\nu)} =$
               $y^{(\nu)}$ and otherwise, set $z^{(\nu)} = \perp$.
         – If $b^{(\nu)} = 1$:
           i. Compute
              $$y^{(\nu)} = \mathcal{D}_0\left(\widetilde{\mathsf{ans}}_{2,1}^{(\nu)}, \ldots, \widetilde{\mathsf{ans}}_{2,k}^{(\nu)}; \mathsf{aux}_2^{(\nu)}\right).$$
           ii. Check whether $\widetilde{\mathsf{ans}}_{1,i}^{(\nu)} = \widetilde{\mathsf{ans}}_{1,j}^{(\nu)}$ holds. If the equality holds, set $z^{(\nu)} =$
               $y^{(\nu)}$ and otherwise, set $z^{(\nu)} = \perp$.
     (c) Add $z^{(\nu)}$ to $\mathcal{L}$.
  3. If $\mathcal{L} = \{z\}$ for some $z \in \{0,1\}$, output $z$. Otherwise, output $\perp$.
```

Fig. 5. The reconstruction algorithm of Π

The $(1-\epsilon)$-correctness of Π_0 implies that

$$y^{(\nu)} = \mathcal{D}_0(\widetilde{\mathsf{ans}}_{1,1}^{(\nu)}, \dots, \widetilde{\mathsf{ans}}_{1,k}^{(\nu)}; \mathsf{aux}_1^{(\nu)}) = a_\tau$$

with probability $1-\epsilon$.

If the client chooses $(i,j) \in \mathcal{I}$ at Step 1(b) of $\mathcal{Q}$, the $(2,i)$-th entry of $\boldsymbol{Q}^{(\nu)}$ is equal to the $(2,j)$-th entry of $\boldsymbol{Q}^{(\nu)}$, which is $(j, \mathsf{que}_{2,j}^{(\nu)})$. Since Servers i and j are honest, it holds that

$$\widetilde{\mathsf{ans}}_{2,i}^{(\nu)} = \mathcal{A}_0(j, \mathsf{que}_{2,j}^{(\nu)}, \boldsymbol{a}) = \widetilde{\mathsf{ans}}_{2,j}^{(\nu)}.$$

Hence, at Step 2(b), the equality holds with probability 1.

Therefore, $z^{(\nu)}$ is always set to $y^{(\nu)} \in \{0,1\}$, which is equal to a_τ with probability $1-\epsilon$.

Privacy. Observe that a query vector $(\mathsf{que}_i)_{i\in[k]}$ generated by $\mathcal{Q}$ contains nothing more than 2λ independent query vectors $(\mathsf{que}_{m,i}^{(\nu)})_{i\in[k]}$ $(m \in \{1,2\}, \nu \in [\lambda])$, each generated by $\mathcal{Q}_0$. Therefore, the t-privacy of Π follows from that of Π_0.

Error detection. We prove that Π is $(b; 1-\epsilon_{\mathrm{ED}})$-error-detecting. Let $\boldsymbol{a} \in \{0,1\}^n$ and $\tau \in [n]$. Without loss of generality, we may assume that the server S_1 is honest. Let $T = [k] \setminus \{1\}$ and $f \in \mathcal{F}_T^{\Pi}$ be a tampering function for Π with respect to T.

Let $\mathcal{I}_0 = \mathcal{I} \times \{0,1\}$. Let $\mathfrak{R}_{\mathcal{Q}}$ denote the set of all random strings for $\mathcal{Q}$, that is, $\mathcal{I}_0^\lambda \times (\mathfrak{R}_{\mathcal{Q}_0}^N)^\lambda$. We suppose that any $(\pi, r) \in \mathfrak{R}_{\mathcal{Q}}$ is decomposed into $\pi = (i^{(\nu)}, j^{(\nu)}, b^{(\nu)})_{\nu\in[\lambda]}$ and $r = (r_m^{(\nu)})_{m\in\{1,2\},\nu\in[\lambda]}$, where $(i^{(\nu)}, j^{(\nu)}, b^{(\nu)}) \in \mathcal{I}_0$ and $r_m^{(\nu)} \in \mathfrak{R}_{\mathcal{Q}_0}$. We naturally identify any event A with a subset of $\mathfrak{R}_{\mathcal{Q}}$ consisting of all random strings on which A occurs.

Let E denote the event in which $\mathcal{D}_0$ outputs an incorrect value even if all servers are honest. Formally, we define

$$\mathsf{E} = \left\{ (\pi, r) \in \mathfrak{R}_{\mathcal{Q}} : \begin{array}{l} \exists \nu \in [\lambda], \exists m \in \{1,2\}, \\ \mathcal{D}_0((\mathcal{A}_0(i, \mathsf{que}_{m,i}^{(\nu)}, \boldsymbol{a}))_{i\in[k]}; \mathsf{aux}_m^{(\nu)}) = 1 - a_\tau \end{array} \right\},$$

where $((\mathsf{que}_{m,i}^{(\nu)})_{i\in[k]}; \mathsf{aux}_m^{(\nu)}) = \mathcal{Q}_0(\tau; r_m^{(\nu)})$ for any $m \in \{1,2\}, \nu \in [\lambda]$. The $(1-\epsilon)$-correctness of Π_0 implies that E occurs with probability at most $2\lambda\epsilon$. Let

$$\mathfrak{R}_{\mathsf{E}} = \{ r \in (\mathfrak{R}_{\mathcal{Q}_0})^\lambda : \exists \pi \in \mathcal{I}_0^\lambda, (\pi, r) \in \mathsf{E} \}.$$

For any $(\pi, r) \in \mathfrak{R}_{\mathcal{Q}}$, let $w(\pi, r) \in \{0, 1, \bot\}$ denote the value outputted by the client when (π, r) is used to generate queries. Let F denote the set of all (π, r)'s such that $w(\pi, r) = 1 - a_\tau$.

Let R be the random variable representing $r \leftarrow_{\$} (\mathfrak{R}_{\mathcal{Q}_0})^\lambda$. We have that

$$\begin{aligned} \Pr[\mathsf{F}] &= \Pr[\mathsf{E} \cap \mathsf{F}] + \Pr\left[\overline{\mathsf{E}} \cap \mathsf{F}\right] \\ &\leq \Pr[\mathsf{E}] + \sum_{r \in (\mathfrak{R}_{\mathcal{Q}_0})^\lambda} \Pr\left[\overline{\mathsf{E}} \cap \mathsf{F} \;\middle|\; R = r\right] \cdot \Pr[R = r] \\ &\leq 2\lambda\epsilon + \sum_{r \notin \mathfrak{R}_{\mathsf{E}}} \Pr\left[\overline{\mathsf{E}} \cap \mathsf{F} \;\middle|\; R = r\right] \cdot \Pr[R = r]. \end{aligned} \tag{1}$$

Fix $r \notin \mathfrak{R}_{\mathsf{E}}$. For every $\nu \in [\lambda]$, let $\mathsf{F}^{(\nu)}$ be the event conditioned on $R = r$ that $z^{(\nu)} = 1 - a_\tau$ at the ν-th iteration of Step 2 of $\mathcal{D}$. We have that

$$\begin{aligned}\Pr\left[\overline{\mathsf{E}} \cap \mathsf{F} \;\middle|\; R = r\right] &\le \Pr_\pi\left[\mathsf{F}^{(1)} \cap \cdots \cap \mathsf{F}^{(\lambda)}\right] \\ &= \prod_{\nu \in [\lambda]} \Pr_\pi\left[\mathsf{F}^{(\nu)} \;\middle|\; \mathsf{F}^{(1)} \cap \cdots \cap \mathsf{F}^{(\nu-1)}\right] \qquad (2)\end{aligned}$$

Furthermore, we have that

$$\begin{aligned}&\Pr_\pi\left[\mathsf{F}^{(\nu)} \;\middle|\; \mathsf{F}^{(1)} \cap \cdots \cap \mathsf{F}^{(\nu-1)}\right] \\ = &\sum_{\substack{\pi^{(1)},\dots,\pi^{(\nu-1)},\\ \pi^{(\nu+1)},\dots,\pi^{(\lambda)}}} \Pr\left[\pi^{(1)}, \dots, \pi^{(\nu-1)}, \pi^{(\nu+1)}, \dots, \pi^{(\lambda)}\right] \\ &\times \Pr_{\pi^{(\nu)}}\left[\mathsf{F}^{(\nu)} \;\middle|\; \mathsf{F}^{(1)} \cap \cdots \cap \mathsf{F}^{(\nu-1)}, \pi^{(1)}, \dots, \pi^{(\nu-1)}, \pi^{(\nu+1)}, \dots, \pi^{(\lambda)}\right]. \qquad (3)\end{aligned}$$

Fix $\pi^{(1)}, \dots, \pi^{(\nu-1)}, \pi^{(\nu+1)}, \dots, \pi^{(\lambda)} \in \mathcal{I}_0$. For ease of reading, let COND denote the condition of the probability (3). Define an event BAD that the client picks $\pi^{(\nu)} = (i^{(\nu)}, j^{(\nu)}, b^{(\nu)}) \in \mathcal{I}_0$ such that $i^{(\nu)} \neq 1$. In other words, BAD means that the client fails to guess that S_1 is honest. Then, we have that

$$\begin{aligned}&\Pr_{\pi^{(\nu)}}\left[\mathsf{F}^{(\nu)} \;\middle|\; \mathsf{COND}\right] \\ &= \Pr[\mathsf{BAD}] \cdot \Pr_{\pi^{(\nu)}}\left[\mathsf{F}^{(\nu)} \;\middle|\; \mathsf{COND}, \mathsf{BAD}\right] \\ &\quad + \Pr\left[\overline{\mathsf{BAD}}\right] \cdot \Pr_{\pi^{(\nu)}}\left[\mathsf{F}^{(\nu)} \;\middle|\; \mathsf{COND}, \overline{\mathsf{BAD}}\right] \\ &\le \frac{2k(k-1) - 2(k-1)}{2k(k-1)} + \frac{2(k-1)}{2k(k-1)} \cdot \Pr_{\pi^{(\nu)}}\left[\mathsf{F}^{(\nu)} \;\middle|\; \mathsf{COND}, \overline{\mathsf{BAD}}\right] \\ &\le \frac{k-1}{k} + \frac{1}{k} \cdot \Pr_{\pi^{(\nu)}}\left[\mathsf{F}^{(\nu)} \;\middle|\; \mathsf{COND}, \overline{\mathsf{BAD}}\right]. \qquad (4)\end{aligned}$$

We will show that

$$\Pr_{\pi^{(\nu)}}\left[\mathsf{F}^{(\nu)} \;\middle|\; \mathsf{COND}, \overline{\mathsf{BAD}}\right] \le 1 - \frac{1}{2(k-1)}. \qquad (5)$$

Let X denote the set of all $\pi^{(\nu)} \in \mathcal{I}_0$ such that

- $\pi^{(\nu)} \in \overline{\mathsf{BAD}}$, i.e., it has the form of $\pi^{(\nu)} = (1, j^{(\nu)}, b^{(\nu)})$;
- COND occurs on $\pi^{(\nu)}$, i.e., it holds that

$$\pi := (\pi^{(1)}, \dots, \pi^{(\nu-1)}, \pi^{(\nu)}, \pi^{(\nu+1)}, \dots, \pi^{(\lambda)}) \in \mathsf{F}^{(1)} \cap \cdots \cap \mathsf{F}^{(\nu-1)}.$$

Let Y denote a subset consisting of all $\pi^{(\nu)} \in X$ satisfying $\pi \in \mathsf{F}^{(1)} \cap \cdots \cap \mathsf{F}^{(\nu-1)} \cap \mathsf{F}^{(\nu)}$.

If $X = \emptyset$, then (5) clearly holds. If $X \neq \emptyset$, choose $\pi^{(\nu)} = (1, j^{(\nu)}, b^{(\nu)}) \in X$ arbitrarily. Denote the queries sent to the malicious servers $\mathsf{S}_2, \ldots, \mathsf{S}_k$ when $\pi^{(\nu)}$ is picked at Step 1(b) of $\mathcal{Q}$, by

$$(2, \mathsf{que}_{1,2}^{(\mu)}), \ldots, (k, \mathsf{que}_{1,k}^{(\mu)}), (2, \mathsf{que}_{2,2}^{(\mu)}), \ldots, (k, \mathsf{que}_{2,k}^{(\mu)}),\ \mu \in [\lambda] \tag{6}$$

We can see that if another $(1, j, b) \in \mathcal{I}_0$ is picked, the queries sent to the malicious servers are the same as (6). Since the tampering function f is deterministic, the answers returned by them are also the same regardless of what is picked as $(j^{(\nu)}, b^{(\nu)})$ at Step 1(b) of $\mathcal{Q}$. Therefore, if COND occurs on $\pi^{(\nu)}$, COND occurs on every $(1, j, b) \in \mathcal{I}_0$. In particular, we have that $|X| = 2(k-1)$ if $X \neq \emptyset$.

We have seen that the answers returned by the malicious servers $\mathsf{S}_2, \ldots, \mathsf{S}_k$ are the same for any $\pi^{(\nu)} \in X$. We denote the answers by

$$\widetilde{\mathsf{ans}}_{1,2}^{(\nu)}, \ldots, \widetilde{\mathsf{ans}}_{1,k}^{(\nu)}, \widetilde{\mathsf{ans}}_{2,2}^{(\nu)}, \ldots, \widetilde{\mathsf{ans}}_{2,k}^{(\nu)}.$$

If all of them are correct, i.e., $\widetilde{\mathsf{ans}}_{m,j}^{(\nu)} = \mathcal{A}_0(j, \mathsf{que}_{m,j}^{(\nu)}, \boldsymbol{a})$, we obtain that $Y = \emptyset$. This is because at one of the rows of $\boldsymbol{A}^{(\nu)}$, the client computes

$$y^{(\nu)} = \mathcal{D}_0(\widetilde{\mathsf{ans}}_{m,1}^{(\nu)}, \ldots, \widetilde{\mathsf{ans}}_{m,k}^{(\nu)}; \mathsf{aux}_m^{(\nu)}).$$

Since we assume E does not occur, i.e., $r \notin \mathfrak{R}_{\mathsf{E}}$, an outcome of $\mathcal{D}_0$ never results in $1 - a_\tau$ and hence $\mathsf{F}^{(\nu)}$ never occurs. Assume that there exist $m \in \{1, 2\}$ and $j \in \{2, 3, \ldots, k\}$ such that

$$\widetilde{\mathsf{ans}}_{m,j}^{(\nu)} \neq \mathcal{A}_0(j, \mathsf{que}_{m,j}^{(\nu)}, \boldsymbol{a}).$$

We can see that $(1, j, 1) \notin Y$ if $m = 1$, and that $(1, j, 0) \notin Y$ if $m = 2$. To see this, consider the case of $m = 1$. If $j^{(\nu)} = j$ and $b^{(\nu)} = 1$ are picked at Step 1(b) of $\mathcal{Q}$, the client detects errors (i.e., outputs $\perp$) since he finds the inconsistency

$$\widetilde{\mathsf{ans}}_{1,1}^{(\nu)} = \mathcal{A}_0(j, \mathsf{que}_{1,j}^{(\nu)}, \boldsymbol{a}) \neq \widetilde{\mathsf{ans}}_{1,j}^{(\nu)}.$$

The other case of $m = 2$ is similar. Therefore, if $X \neq \emptyset$,

$$\Pr_{\pi^{(\nu)}}\left[\mathsf{F}^{(\nu)} \,\middle|\, \mathsf{COND}, \overline{\mathsf{BAD}}\right] = \frac{|Y|}{|X|} \leq \frac{2(k-1)-1}{2(k-1)} = 1 - \frac{1}{2(k-1)},$$

which implies (5).

Finally, we obtain from (3) and (4) that

$$\begin{aligned}
&\Pr_\pi\left[\mathsf{F}^{(\nu)} \,\middle|\, \mathsf{F}^{(1)} \cap \cdots \cap \mathsf{F}^{(\nu-1)}\right] \\
&\leq \sum_{\substack{\pi^{(1)}, \ldots, \pi^{(\nu-1)}, \\ \pi^{(\nu+1)}, \ldots, \pi^{(\lambda)}}} \Pr\left[\pi^{(1)}, \ldots, \pi^{(\nu-1)}, \pi^{(\nu+1)}, \ldots, \pi^{(\lambda)}\right] \\
&\quad \times \left(\frac{k-1}{k} + \frac{1}{k} \cdot \left(1 - \frac{1}{2(k-1)}\right)\right) \\
&\leq 1 - \frac{1}{2k(k-1)}
\end{aligned}$$

and hence (2) implies that

$$\Pr\left[\overline{\mathsf{E}} \cap \mathsf{F} \mid R = r\right] \leq \left(1 - \frac{1}{2k(k-1)}\right)^{\lambda}.$$

Therefore, the $(b; 1-\epsilon_{\mathrm{ED}})$-error detection follows from (1) and

$$\Pr[\mathsf{F}] \leq 2\lambda\epsilon + \left(1 - \frac{1}{2k(k-1)}\right)^{\lambda} = \epsilon_{\mathrm{ED}}.$$

C Proof of Lemma 1

Define Π as follows:

- Iterate Π_0 λ times in parallel.
- Let $y_i \in \{0,1\}$ be the output of the i-th iteration of Π_0 for $i \in [\lambda]$. If there exists $y \in \{0,1\}$ such that $|\{i : y_i = y\}| > |\{i : y_i = 1 - y\}|$, output y. Otherwise, output 0.

Clearly, the communication complexity of Π is λ times larger than that of Π_0 and the t-privacy of Π directly follows from that of Π_0. Let $\boldsymbol{a} \in \{0,1\}^n$ be a database and $\tau \in [n]$ be a client's index. The outputs of Π_0 are independent and each output is equal to a_τ with probability $1 - \epsilon_0$.

Let X_i be a random variable over $\{0,1\}$ defined as $X_i = 1$ if and only if $y_i = a_\tau$. $X_1, \ldots, X_\lambda$ are i.i.d. random variables such that $p = \mathbb{E}[X_1] = 1 - \epsilon_0$. It then follows from the Chernoff bound that

$$\begin{aligned}
\Pr[\Pi \text{ outputs } 1 - a_\tau] &\leq \Pr\left[\frac{1}{\lambda}\sum_{i=1}^{\lambda} X_i \leq \frac{1}{2}\right] \\
&= \Pr\left[\frac{1}{\lambda}\sum_{i=1}^{\lambda} X_i \leq p + \left(\frac{1}{2} - p\right)\right] \\
&\leq \left(\left(\frac{p}{1/2}\right)^{1/2}\left(\frac{1-p}{1/2}\right)^{1/2}\right)^{\lambda} \\
&= (4p(1-p))^{\lambda/2} \\
&= (2\sqrt{\epsilon_0(1-\epsilon_0)})^{\lambda} \\
&\leq \exp\left(-2\left(\frac{1}{2} - \epsilon_0\right)^2 \lambda\right).
\end{aligned}$$

The last inequality follows from

$$\frac{1}{2}\ln(4\epsilon_0(1-\epsilon_0)) = \frac{1}{2}\ln(1 - 4x^2) \leq -2x^2,$$

where $x = 1/2 - \epsilon_0$.

D Proof of Theorem 3

For $\mathit{\Pi}_0 = (\mathcal{Q}_0, \mathcal{A}_0, \mathcal{D}_0)$, we consider a PIR scheme $\mathit{\Pi} = (\mathcal{Q}, \mathcal{A}, \mathcal{D})$ where $(\mathcal{Q}, \mathcal{A})$ runs N independent instances of $(\mathcal{Q}_0, \mathcal{A}_0)$ between a client and every subset of k servers and $\mathcal{D}$ is defined as follows: For each of N executions of $(\mathcal{Q}_0, \mathcal{A}_0)$, $\mathcal{D}$ runs $\mathcal{D}_0$ on the corresponding input and adds the output to a list $\mathcal{L}$. If $\mathcal{L} = \{s\}$ or $\mathcal{L} = \{s, \perp\}$ for some $s \in \{0,1\}$, then $\mathcal{D}$ outputs s and otherwise outputs 0. The communication complexity of $\mathit{\Pi}$ is Nc. Since each execution of $\mathcal{Q}_0$ is done independently, $\mathit{\Pi}$ is also t-private.

We prove that $\mathit{\Pi}$ is $(b; 1-\epsilon_{\mathrm{EC}})$-error-correcting for $\epsilon_{\mathrm{EC}} = N\epsilon_{\mathrm{ED}}$. Let $\boldsymbol{a} \in \{0,1\}^n$ and $\tau \in [n]$. Let $H \in \binom{[\ell]}{k}$ be a set of honest servers. Let $f \in \mathcal{F}^{\mathit{\Pi}}_{\overline{H}}$ be a tampering function for $\mathit{\Pi}$ with respect to $\overline{H}$.

Let $A_1, \ldots, A_N$ be all k-sized subsets of $[\ell]$ such that $A_1 = H$. Let $\mathit{\Pi}_0^{(j)}$ denote the instance of $\mathit{\Pi}_0$ executed by the client and servers in A_j. During the execution of $\mathit{\Pi}_0^{(j)}$, the client generates

$$\mathcal{Q}_0(\tau; r_j) = ((\mathsf{que}_i^{(j)})_{i \in A_j}; \mathsf{aux}^{(j)}),$$

where $r_j \in \mathfrak{R}_{\mathcal{Q}_0}$ and $\mathsf{que}_i^{(j)}$ is sent to S_i. Then, S_i receives

$$\mathsf{que}'_i = \{\mathsf{que}_i^{(j)} : j \in [N] \text{ with } i \in A_j\}.$$

In $\mathit{\Pi}_0^{(1)}$, for any $i \in A_1$, S_i returns

$$\widetilde{\mathsf{ans}}_i^{(1)} = \mathsf{ans}_i^{(1)} = \mathcal{A}_0(i, \mathsf{que}_i^{(1)}, \boldsymbol{a}).$$

In each $\mathit{\Pi}_0^{(j)}$ for $j \neq 1$, any server S_i in A_j returns

$$\widetilde{\mathsf{ans}}_i^{(j)} = \begin{cases} \mathsf{ans}_i^{(j)} = \mathcal{A}_0(i, \mathsf{que}_i^{(j)}, \boldsymbol{a}), & \text{if } i \in H, \\ f_i^{(j)}(\{\mathsf{que}'_{i'}\}_{i' \in \overline{H}}, \boldsymbol{a}), & \text{otherwise,} \end{cases}$$

where $f_i^{(j)}$ is a function determined by f. It then follows from our definition of $\mathcal{D}$ that

$$\begin{aligned} &\Pr[\mathcal{D} \text{ outputs } y \neq a_\tau] \\ &\leq \Pr\left[\mathcal{D}_0((\mathsf{ans}_i^{(1)})_{i \in H}; \mathsf{aux}^{(1)}) \neq a_\tau\right] \\ &\quad + \sum_{j=2}^{N} \Pr\left[\mathcal{D}_0((\mathsf{ans}_i^{(j)})_{i \in H \cap A_j}, (\widetilde{\mathsf{ans}}_i^{(j)})_{i \in \overline{H} \cap A_j}; \mathsf{aux}^{(j)}) \notin \{a_\tau, \perp\}\right] \\ &\leq \epsilon_{\mathrm{ED}} + \sum_{j=2}^{N} \Pr\left[\mathcal{D}_0((\mathsf{ans}_i^{(j)})_{i \in H \cap A_j}, (\widetilde{\mathsf{ans}}_i^{(j)})_{i \in \overline{H} \cap A_j}; \mathsf{aux}^{(j)}) \notin \{a_\tau, \perp\}\right]. \end{aligned}$$

Therefore it is enough to show that

$$p_0 := \Pr\left[\mathcal{D}_0((\mathsf{ans}_i^{(j)})_{i \in H \cap A_j}, (\widetilde{\mathsf{ans}}_i^{(j)})_{i \in \overline{H} \cap A_j}; \mathsf{aux}^{(j)}) \notin \{a_\tau, \perp\}\right] \leq \epsilon_{\mathrm{ED}}$$

for any $j \in [N] \setminus \{1\}$.

Let $j \in [N] \setminus \{1\}$. Fix $r_{-j} = (r_m)_{m \in [N] \setminus \{j\}}$ arbitrarily. Then $\mathsf{que}_i^{(m)}$ is a fixed constant for any $m \in [N] \setminus \{j\}$ and $i \in A_m$. Therefore for $i \in A_j$, we can write

$$\widetilde{\mathsf{ans}}_i^{(j)} = f_i^{(j)}(\{\mathsf{que}'_{i'}\}_{i' \in \overline{H}}, \boldsymbol{a}) = g_{i,r_{-j}}(\{\mathsf{que}_{i'}^{(j)}\}_{i' \in \overline{H} \cap S_j}, \boldsymbol{a})$$

using some function $g_{i,r_{-j}}$. Let $\mathcal{X}_{-j}$ denote the random variable which represents r_{-j}. Since $|\overline{H} \cap A_j| \leq b$ and Π_0 is $(b, t; 1 - \epsilon_{\mathrm{ED}})$-error-detecting, we have that

$$\begin{aligned} p_0 &= \Pr_{r_j, r_{-j}}\left[\mathcal{D}_0((\mathsf{ans}_i^{(j)})_{i \in H \cap A_j}, (\widetilde{\mathsf{ans}}_i^{(j)})_{i \in \overline{H} \cap A_j}; \mathsf{aux}^{(j)}) \notin \{a_\tau, \bot\}\right] \\ &= \sum_{r_{-j}} \Pr[\mathcal{X}_{-j} = r_{-j}] \Pr_{r_j}\left[\mathcal{D}_0((\mathsf{ans}_i^{(j)})_{i \in H \cap A_j}, (\widetilde{\mathsf{ans}}_i^{(j)})_{i \in \overline{H} \cap A_j}; \mathsf{aux}^{(j)}) \notin \{a_\tau, \bot\}\right] \\ &\leq \sum_{r_{-j}} \Pr[\mathcal{X}_{-j} = r_{-j}] \times \epsilon_{\mathrm{ED}} \\ &= \epsilon_{\mathrm{ED}}. \end{aligned}$$

E Proof of Theorem 5

For $\boldsymbol{a} = (a_1, \ldots, a_n) \in \{0,1\}^n$, define $\boldsymbol{a}^* = (a_1^*, \ldots, a_n^*) \in \{0,1\}^n$ as the same database as $\boldsymbol{a}$ except that $a_1^* = 1 - a_1$. Let $B \subseteq [\ell]$ be a subset of size b and let $B' = [\ell] \setminus B$. Let f be a tampering function such that $f(\mathsf{que}_1, \ldots, \mathsf{que}_\ell, \boldsymbol{a}) = (\widetilde{\mathsf{ans}}_i)_{i \in [\ell]}$, where

$$\widetilde{\mathsf{ans}}_i = \begin{cases} \mathcal{A}(i, \mathsf{que}_i, \boldsymbol{a}), & \text{if } i \notin B, \\ \mathcal{A}(i, \mathsf{que}_i, \boldsymbol{a}^*), & \text{if } i \in B. \end{cases}$$

for any $i \in [\ell]$ and $\mathsf{que}_i \in \{0,1\}^{c_{\mathsf{que}}}$. Also, let f' be a tampering function such that $f'(\mathsf{que}_1, \ldots, \mathsf{que}_\ell, \boldsymbol{a}) = (\widetilde{\mathsf{ans}}'_i)_{i \in [\ell]}$, where

$$\widetilde{\mathsf{ans}}'_i = \begin{cases} \mathcal{A}(i, \mathsf{que}_i, \boldsymbol{a}), & \text{if } i \notin B', \\ \mathcal{A}(i, \mathsf{que}_i, \boldsymbol{a}^*), & \text{if } i \in B' \end{cases}$$

for any $i \in [\ell]$ and $\mathsf{que}_i \in \{0,1\}^{c_{\mathsf{que}}}$. Note that $f \in \mathcal{F}_B^\Pi$ and $f' \in \mathcal{F}_{B'}^\Pi$. Also note that $(\boldsymbol{a}^*)^* = \boldsymbol{a}$ and that $i \notin B$ is equivalent to $i \in B'$. Thus, we have that

$$f(\mathsf{que}_1, \ldots, \mathsf{que}_\ell, \boldsymbol{a}) = f'(\mathsf{que}_1, \ldots, \mathsf{que}_\ell, \boldsymbol{a}^*) \tag{7}$$

for any $\boldsymbol{a} \in \{0,1\}^n$ and $\mathsf{que}_1, \ldots, \mathsf{que}_\ell \in \{0,1\}^{c_{\mathsf{que}}}$.

Fix $\boldsymbol{a} \in \{0,1\}^n$ arbitrarily. Define a subset S (resp. S') of $\mathfrak{R}_\mathcal{Q}$ as

$$\begin{aligned} S &= \{r \in \mathfrak{R}_\mathcal{Q} : \mathcal{D}(f(\mathsf{que}_1, \ldots, \mathsf{que}_\ell, \boldsymbol{a}); \mathsf{aux}) = a_1\}, \\ S' &= \{r \in \mathfrak{R}_\mathcal{Q} : \mathcal{D}(f'(\mathsf{que}_1, \ldots, \mathsf{que}_\ell, \boldsymbol{a}^*); \mathsf{aux}) = a_1^*\}, \end{aligned}$$

where $(\mathsf{que}_1, \ldots, \mathsf{que}_\ell; \mathsf{aux}) = \mathcal{Q}(1; r)$. It follows from Eq. (7) and $a_1^* = 1 - a_1 \neq a_1$ that $S \cap S' = \emptyset$. On the other hand, since $|B| = b$ and $|B'| = \ell - b \leq b$, the $(b; 1-\epsilon_{\mathrm{EC}})$-error correction of Π implies that $|S| \geq (1-\epsilon_{\mathrm{EC}})|\mathfrak{R}_{\mathcal{Q}}|$ and $|S'| \geq (1-\epsilon_{\mathrm{EC}})|\mathfrak{R}_{\mathcal{Q}}|$. Therefore, we have that

$$|\mathfrak{R}_{\mathcal{Q}}| \geq |S \cup S'| = |S| + |S'| \geq 2(1-\epsilon_{\mathrm{EC}})|\mathfrak{R}_{\mathcal{Q}}|$$

and $\epsilon_{\mathrm{EC}} \geq 1/2$.

F Proof of Theorem 7

Setup. Given $(\xi_k)_{k\in[M]} \in \mathbb{F}_p^M$ and $\alpha \in \mathbb{F}_q \setminus \{0\}$, C does the following:

1. For each $k \in [M]$:
   (a) He chooses $\beta_k \leftarrow\$\, \mathbb{F}_q$.
   (b) He sets $T_k(X) = \xi_k + \beta_k X \in \mathbb{F}_q[X]$.
   (c) He sets $\rho_k = T_k(\alpha) \in \mathbb{F}_q$.
2. He outputs $\mathsf{msg} = (\xi_k)_{k\in[M]}$, $\mathsf{tag} = (T_k(X))_{k\in[M]}$ and $\mathsf{key} = (\rho_k)_{k\in[M]}$.

Evaluation (M). Given an M-variate polynomial $F = F(X_1, \dots, X_M)$ of total degree at most d over $\mathbb{F}_p$, $\mathsf{msg} \in \mathbb{F}_p^M$ and $\mathsf{tag} \in (\mathbb{F}_q[X])^M$, M outputs $\gamma = F(\mathsf{msg})$ and $G(X) = F(\mathsf{tag})$.

Evaluation (H). Given the polynomial $F = F(X_1, \dots, X_M)$ and $\mathsf{key} \in \mathbb{F}_q^M$, H outputs $\sigma = F(\mathsf{key})$.

Verification. Given $\alpha \in \mathbb{F}_q \setminus \{0\}$, $\gamma \in \mathbb{F}_p$, $G(X) \in \mathbb{F}_q[X]$ and $\sigma \in \mathbb{F}_q$, C outputs $\mathsf{detect} = 1$ if $\gamma = G(0)$ and $\sigma = G(\alpha)$ and otherwise, outputs $\mathsf{detect} = 0$.

Fig. 6. A message authentication protocol Σ^{MAC}

Let $\lambda \in \mathbb{N}$ be such that

$$q := p^\lambda > N\ell^2 d \left(\epsilon - \binom{\ell}{b}\epsilon_0\right)^{-1}. \tag{8}$$

Let Σ^{MAC} be a message authentication protocol shown in Fig. 6 where C verifies that a server M correctly computes a polynomial F with help of another server H. Let Π be an ℓ-server PIR scheme $\Pi = (\mathcal{Q}, \mathcal{A}, \mathcal{D})$ described in Fig. 7.

Communication complexity. Observe that $\mathsf{msg}_i = \mathsf{que}_i^{(0)}$ is an M-dimensional vector over $\mathbb{F}_p$, each tag_{ij} consists of M polynomials of degree 1 over $\mathbb{F}_q$, and key_{ji} is an M-dimensional vector over $\mathbb{F}_q$. Also observe that $\gamma_i^{(\mu)}$ is a field element of $\mathbb{F}_p$, $G_{ij}(X)$ is a polynomial of degree d over $\mathbb{F}_q$, and $\sigma_{ji}^{(\mu)}$ is a field element of $\mathbb{F}_q$. Since q can be chosen arbitrarily as long as (8) holds, the communication complexity of Π is

$$\begin{aligned} &\ell\left(O(M\log p + M\ell\log q) + O(N\log p + N\ell d\log q)\right) \\ &= O\left(\ell^2(M+Nd)\log\frac{Nd\ell}{\epsilon - \binom{\ell}{b}\epsilon_0}\right). \end{aligned}$$

```
Notations.
    – A $(\ell - b, \ell; 1 - \epsilon_0)$-robust $\ell$-server PIR scheme $\Pi_0 = (\mathcal{Q}_0, \mathcal{A}_0, \mathcal{D}_0)$ with the following properties:
        • For any $i \in [\ell]$, a query $\mathsf{que}_i^{(0)}$ is an $M$-dimensional vector over $\mathbb{F}_p$;
        • For any $\boldsymbol{a} \in \{0,1\}^n$ and any $i \in [\ell]$, $\mathcal{A}_0(i, \cdot, \boldsymbol{a})$ is a tuple $(F_{\boldsymbol{a},i}^{(\mu)})_{\mu \in [N]}$ of $M$-variate polynomials of total degree at most $d$ over $\mathbb{F}_p$.
    – A field extension $\mathbb{F}_q / \mathbb{F}_p$ of degree $\lambda$
    – A message authentication protocol $\Sigma^{\mathsf{MAC}}$

$\mathcal{Q}(\tau)$. Given $\tau \in [n]$:
    1. Choose $r^{(0)} \leftarrow_{\$} \mathfrak{R}_{\mathcal{Q}_0}$.
    2. Let $(\mathsf{que}_1^{(0)}, \ldots, \mathsf{que}_\ell^{(0)}; \mathsf{aux}^{(0)}) = \mathcal{Q}_0(\tau; r^{(0)})$.
    3. For each $i, j \in [\ell]$, do the following:
        (a) Choose $\alpha_{ij} \leftarrow_{\$} \mathbb{F}_q \setminus \{0\}$.
        (b) Execute "Setup" in $\Sigma^{\mathsf{MAC}}$ on input $\mathsf{que}_i^{(0)} \in \mathbb{F}_p^M$ and $\alpha_{ij}$.
        (c) Let $\mathsf{msg}_i := \mathsf{msg}$, $\mathsf{tag}_{ij} := \mathsf{tag}$ and $\mathsf{key}_{ij} := \mathsf{key}$ be the outputs.
    4. For each $i \in [\ell]$, let $\mathsf{que}_i = (\mathsf{msg}_i, (\mathsf{tag}_{ij})_{j \in [\ell]}, (\mathsf{key}_{ji})_{j \in [\ell]})$.
    5. Let $\mathsf{aux} = (\mathsf{aux}^{(0)}, (\alpha_{ij})_{i,j \in [\ell]})$.
    6. Output $(\mathsf{que}_1, \ldots, \mathsf{que}_\ell; \mathsf{aux})$.

$\mathcal{A}(i, \mathsf{que}_i, \boldsymbol{a})$. Given $i \in [\ell]$, $\mathsf{que}_i = (\mathsf{msg}_i, (\mathsf{tag}_{ij})_{j \in [\ell]}, (\mathsf{key}_{ji})_{j \in [\ell]})$ and $\boldsymbol{a} \in \{0,1\}^n$:
    1. For each $j \in [\ell]$:
        (a) For each $\mu \in [N]$:
            i. Execute "Evaluation (M)" in $\Sigma^{\mathsf{MAC}}$ on input $F_{\boldsymbol{a},i}^{(\mu)}$, $\mathsf{msg}_i$ and $\mathsf{tag}_{ij}$.
            ii. Let $\gamma_i^{(\mu)} := \gamma$ and $G_{ij}^{(\mu)}(X) := G(X)$ be the outputs.
    2. For each $j \in [\ell]$:
        (a) For each $\mu \in [N]$:
            i. Execute "Evaluation (H)" in $\Sigma^{\mathsf{MAC}}$ on input $F_{\boldsymbol{a},j}^{(\mu)}$ and $\mathsf{key}_{ji}$.
            ii. Let $\sigma_{ji}^{(\mu)} := \sigma$ be the output.
    3. Output
        $$\mathsf{ans}_i = \left\{ \boldsymbol{y}_i := \left(\gamma_i^{(\mu)}\right)_{\mu \in [N]}, \left(G_{ij}^{(\mu)}(X), \sigma_{ji}^{(\mu)}\right)_{j \in [\ell], \mu \in [N]} \right\}.$$

$\mathcal{D}(\widetilde{\mathsf{ans}}_1, \ldots, \widetilde{\mathsf{ans}}_\ell; \mathsf{aux})$. Given $\widetilde{\mathsf{ans}}_i$ $(i \in [\ell])$ and $\mathsf{aux} = (\mathsf{aux}^{(0)}, (\alpha_{ij})_{i,j \in [\ell]})$:
    1. For each $i \in [\ell]$, parse $\widetilde{\mathsf{ans}}_i$ as $\widetilde{\mathsf{ans}}_i = (\widetilde{\boldsymbol{y}}_i, (\widetilde{G}_{ij}^{(\mu)}(X), \widetilde{\sigma}_{ji}^{(\mu)})_{j \in [\ell], \mu \in [N]})$. and parse $\widetilde{\boldsymbol{y}}_i$ as $\widetilde{\boldsymbol{y}}_i = (\widetilde{\gamma}_i^{(\mu)})_{\mu \in [N]}$.
    2. For each $i, j \in [\ell]$, do the following:
        (a) Execute "Verification" in $\Sigma^{\mathsf{MAC}}$ on input $\alpha_{ij}$, $\widetilde{\gamma}_i^{(\mu)}$, $\widetilde{G}_{ij}^{(\mu)}(X)$ and $\widetilde{\sigma}_{ij}^{(\mu)}$.
        (b) Let $\mathsf{detect}_{ij} := \mathsf{detect}$ be the output.
    3. Let $A$ be any subset of $\{i \in [\ell] : \sum_{j \in [\ell]} \mathsf{detect}_{ij} \le b\}$ of size $\ell - b$.
    4. Output $\mathcal{D}_0(A, (\widetilde{\boldsymbol{y}}_i)_{i \in A}; \mathsf{aux}^{(0)})$.
```

Fig. 7. An error-correcting PIR scheme Π in Theorem 7

Privacy. At Step 3(c) of $\mathcal{Q}$, we can write

$$\begin{aligned}\mathsf{msg}_i &= \mathsf{que}_i^{(0)} = (\xi_{ik})_{k\in[M]} \in \mathbb{F}_p^M,\\ \mathsf{tag}_{ij} &= (T_{ijk}(X))_{k\in[M]} \in (\mathbb{F}_q[X])^M,\\ \mathsf{key}_{ij} &= (\rho_{ijk})_{k\in[M]} \in \mathbb{F}_q^M.\end{aligned}$$

Then, $T_{ijk}(X)$ can be written as $T_{ijk}(X) = \xi_{ik} + \beta_{ijk}X$ for a field element $\beta_{ijk} \in \mathbb{F}_q$.

Let $B \in \binom{[\ell]}{t}$. A query for S_i generated by $\mathcal{Q}(\tau)$ is

$$\mathsf{que}_i = (\mathsf{que}_i^{(0)}, (T_{ijk}(X), \rho_{jik})_{j\in[\ell],k\in[M]}).$$

In view of the t-privacy of Π_0, it is sufficient to show that $\{\mathsf{que}_i : i \in B\}$ can be simulated from $\{\mathsf{que}_i^{(0)} : i \in B\}$. Observe that

$$T_{ijk}(X) = \xi_{ik} + \beta_{ijk}X,$$

where $\beta_{ijk} \leftarrow_\$ \mathbb{F}_q$. Since the β_{ijk}'s are independent of any query, $(T_{ijk}(X))_{j\in[\ell],k\in[M]}$ can be simulated from $\mathsf{que}_i^{(0)}$. Observe that

$$\rho_{jik} = T_{jik}(\alpha_{ji}) = \xi_{jk} + \beta_{jik}\alpha_{ji}.$$

If $j \in \overline{B}$ and $k \in [M]$, ρ_{jik} is a uniformly random element independent of ξ_{jk} since $\beta_{jik} \leftarrow_\$ \mathbb{F}_q$ is unknown to B and $\alpha_{ji} \neq 0$. If $j \in B$ and $k \in [M]$, ρ_{jik} can be simulated by choosing $\alpha_{ji} \leftarrow_\$ \mathbb{F}_q \setminus \{0\}$ and setting $\rho_{jik} = T_{jik}(\alpha_{ji})$ since S_j knows T_{jik}. Thus, $\{(\rho_{jik})_{j\in[\ell],k\in[M]} : i \in B\}$ contains nothing more than $\{\mathsf{que}_i^{(0)} : i \in B\}$.

Error correction. Let $\boldsymbol{a} \in \{0,1\}^n$ and $\tau \in [n]$. Let $H \in \binom{[\ell]}{k}$ be a set of honest servers. Let $f \in \mathcal{F}^{\Pi}_{\overline{H}}$ be a tampering function for Π with respect to $\overline{H}$.

First, we show that $\{i \in [\ell] : \sum_{j\in[\ell]} v_{ij} \leq b\}$ contains at least $\ell - b$ elements in order to guarantee the existence of a set A at Step 2. It is sufficient to show that

$$H \subseteq \{i \in [\ell] : \sum_{j\in[\ell]} \mathsf{detect}_{ij} \leq b\}. \tag{9}$$

Indeed, let $i, j \in H$. At Step 1 of $\mathcal{D}$, it holds that for any $\mu \in [N]$,

$$\widetilde{\gamma}_i^{(\mu)} = \gamma_i^{(\mu)} = F_{\boldsymbol{a},i}^{(\mu)}(\mathsf{que}_i^{(0)}) = F_{\boldsymbol{a},i}^{(\mu)}((T_{ijk}(0))_{k\in[M]}) = G_{ij}(0) = \widetilde{G}_{ij}(0)$$

and

$$\widetilde{\sigma}_{ij}^{(\mu)} = \sigma_{ij}^{(\mu)} = F_{\boldsymbol{a},i}^{(\mu)}((\rho_{ijk})_{k\in[M]}) = F_{\boldsymbol{a},i}^{(\mu)}((T_{ijk}(\alpha_{ij}))_{k\in[M]}) = G_{ij}(\alpha_{ij}) = \widetilde{G}_{ij}(\alpha_{ij}).$$

We have that $\mathsf{detect}_{ij} = 0$ and hence

$$\sum_{j\in[\ell]} \mathsf{detect}_{ij} = \sum_{j\in\overline{H}} \mathsf{detect}_{ij} \leq b$$

for all $i \in H$. Thus, (9) holds.

Let $r \in \mathfrak{R}_{\mathcal{Q}}$ be any random string for $\mathcal{Q}$. Observe that $r \in \mathfrak{R}_{\mathcal{Q}}$ consists of $r^{(0)} \in \mathfrak{R}_{\mathcal{Q}_0}$ and $\alpha_{ij}, \beta_{ijk} \in \mathbb{F}_q$ for $i, j \in [\ell], k \in [M]$. We note that all items obtained during the execution of Π are functions of r. Let

$$\mathcal{Q}(\tau; r) = (\mathsf{que}_1(r), \ldots, \mathsf{que}_\ell(r); \mathsf{aux}(r)),$$
$$f(\mathsf{que}_1(r), \ldots, \mathsf{que}_\ell(r), \boldsymbol{a}) = (\widetilde{\mathsf{ans}}_1(r), \ldots, \widetilde{\mathsf{ans}}_\ell(r)).$$

Let $A(r)$ be a set constructed at Step 3 of $\mathcal{D}$. Let

$$\boldsymbol{y}_i(r) = (\gamma_i^{(\mu)}(r))_{\mu \in [N]}, \ \widetilde{\boldsymbol{y}}_i(r) = (\widetilde{\gamma}_i^{(\mu)}(r))_{\mu \in [N]}$$

be vectors obtained at Step 3 of $\mathcal{A}$ and at Step 1 of $\mathcal{D}$, respectively. Let $\mathsf{detect}_{ij}(r)$ be a value computed at Step 2(b) of $\mathcal{D}$.

If $\mathcal{D}$ outputs the incorrect value $1 - a_\tau$, there are two possible cases:

Case 1. An incorrect answer $\widetilde{\boldsymbol{y}}_i(r)$ is contained in $A(r)$ without being detected.
Case 2. Every answer $\widetilde{\boldsymbol{y}}_i(r)$ in $A(r)$ is correct but $\mathcal{D}$ outputs $1 - a_\tau$.

Let E denote the event that Case 1 occurs. That is,

$$\mathsf{E} = \{r \in \mathfrak{R}_{\mathcal{Q}} : \exists i \in A(r), \widetilde{\boldsymbol{y}}_i(r) \neq \boldsymbol{y}_i(r)\}.$$

Here, we naturally identify any event with a subset of random strings on which the event occurs. Also, let F denote Case 2 occurs. That is,

$$\mathsf{F} = \overline{\mathsf{E}} \cap \left\{ r \in \mathfrak{R}_{\mathcal{Q}} : \mathcal{D}_0 \left(A(r), (\widetilde{\boldsymbol{y}}_i(r))_{i \in A(r)}; \mathsf{aux}^{(0)}(r) \right) = 1 - a_\tau \right\},$$

where $\mathsf{aux}^{(0)}(r)$ corresponds to $\mathsf{aux}^{(0)}$ in $\mathsf{aux}(r)$. Thus, the probability that $\mathcal{D}$ outputs $1 - a_\tau$ is at most

$$\begin{aligned} &\Pr[\mathsf{E} \cup \mathsf{F}] \\ &\leq \Pr[\mathsf{E}] + \Pr\left[r \leftarrow_{\$} \mathfrak{R}_{\mathcal{Q}} : \exists A \in \binom{[\ell]}{\ell - b}, \mathcal{D}_0 \left(A, (\boldsymbol{y}_i(r))_{i \in A}; \mathsf{aux}^{(0)}(r) \right) = 1 - a_\tau \right] \\ &\leq \Pr[\mathsf{E}] + \binom{\ell}{b} \epsilon_0. \end{aligned} \tag{10}$$

We will show that

$$\Pr[\mathsf{E}] \leq \frac{Nb(\ell - b)d}{q}. \tag{11}$$

Let $i \in \overline{H}$ and $j \in H$. We denote by α_{ij} a component of $r \in \mathfrak{R}_{\mathcal{Q}}$ corresponding to a random field element chosen at Step 3(a) of $\mathcal{Q}$. Let r_{ij} denote all components of r excluding α_{ij}, that is, $r = (\alpha_{ij}, r_{ij})$. Observe that the queries for malicious servers $\{\mathsf{que}_{i'}(r) : i' \in \overline{H}\}$ depend only on r_{ij} and are independent of α_{ij}. Since S_i generates her incorrect answer based only on queries $\mathsf{que}_{i'}(r)$ for $i' \in \overline{H}$,

$$\Delta_{ij}^{(\mu)}(X) := \widetilde{G}_{ij}^{(\mu)}(X) - G_{ij}^{(\mu)}(X)$$

is also determined only by r_{ij}. Define

$$\begin{aligned} R_{ij}^{(\mu)} &= \left\{ r_{ij} : \varDelta_{ij}^{(\mu)}(0) \neq 0 \right\}, \\ S_{ij}^{(\mu)} &= \left\{ (\alpha_{ij}, r_{ij}) \in \mathfrak{R}_{\mathcal{Q}} : r_{ij} \in R_{ij}^{(\mu)} \wedge \varDelta_{ij}^{(\mu)}(\alpha_{ij}) = 0 \right\}. \end{aligned}$$

For any $r_{ij} \in R_{ij}^{(\mu)}$, $\varDelta_{ij}^{(\mu)}(X)$ is a non-zero polynomial of degree d and hence the number of $\alpha_{ij} \in \mathbb{F}_q$ such that $(\alpha_{ij}, r_{ij}) \in S_{ij}^{(\mu)}$ is at most d. Therefore, we obtain that

$$|S_{ij}^{(\mu)}| \leq \sum_{r_{ij} \in R_{ij}^{(\mu)}} \left| \{ \alpha_{ij} : (\alpha_{ij}, r_{ij}) \in S_{ij}^{(\mu)} \} \right| \leq d |R_{ij}^{(\mu)}| \leq \frac{d|R|}{q-1}$$

for any $\mu \in [N]$.

Let $r \in \mathsf{E}$. There exists $i \in A(r)$ such that $\widetilde{\boldsymbol{y}}_i(r) \neq \boldsymbol{y}_i(r)$. Let $\mu \in [N]$ be such that $\widetilde{\gamma}_i^{(\mu)}(r) \neq \gamma_i^{(\mu)}(r)$. Since $\widetilde{\boldsymbol{y}}_j(r) = \boldsymbol{y}_j(r)$ for any $j \in H$, we must have that $i \in \overline{H}$. Since $\ell - b > b$, there exists $j \in H$ such that $\mathsf{detect}_{ij}(r) = 0$ and hence $\widetilde{\gamma}_i^{(\mu)}(r) = \widetilde{G}_{ij}(0)$ and $\widetilde{\sigma}_{ij}^{(\mu)} = \widetilde{G}_{ij}(\alpha_{ij})$. Parse r as $r = (\alpha_{ij}, r_{ij})$. Since $\gamma_i^{(\mu)}(r) = G_{\boldsymbol{a},i,j}(0)$, it holds that

$$\varDelta_{ij}^{(\mu)}(0) = \widetilde{\gamma}_i^{(\mu)}(r) - \gamma_i^{(\mu)}(r) \neq 0$$

and hence $r_{ij} \in R_{ij}^{(\mu)}$. Since $j \in H$,

$$\widetilde{G}_{ij}(\alpha_{ij}) = \widetilde{\sigma}_{ij}^{(\mu)} = \sigma_{ij}^{(\mu)} = G_{ij}(\alpha_{ij})$$

and hence $\varDelta_{ij}^{(\mu)}(\alpha_{ij}) = 0$. We obtain that $r \in S_{ij}^{(\mu)}$. Therefore,

$$|S| \leq \sum_{i \in \overline{H}, j \in H, \mu \in [N]} |S_{ij}^{(\mu)}| \leq Nb(\ell - b) \frac{d|R|}{q-1}$$

and hence (11) holds.

It follows from (10) that

$$\Pr[\mathsf{E} \cup \mathsf{F}] \leq \frac{N\ell^2 d}{q-1} + \binom{\ell}{b} \epsilon_0 \leq \epsilon$$

and hence $\varPi$ satisfies $(b, t; 1 - \epsilon)$-error correction.