Cryptanalysis of a PIR Scheme based on Linear Codes over Rings

Luana Kurmann, Svenja Lage and Violetta Weger

Keywords: Private information retrieval; Codes over rings; Privacy.

Introduction

A Private Information Retrieval (PIR) scheme allows users to retrieve files from a database such that the server does not learn anything about the requested file index. There exist several information-theoretic secure PIR schemes on multiple non-colluding servers. However, if only one server is used, or if we assume that the servers are colluding, we enter the realm of computational PIR (cPIR). While most of the existing cPIR schemes are based on classic number-theoretic problems, such as integer factorization [1, 2, 3, 4], the recent advances in quantum computation require novel and quantum-secure cPIR solutions. Most of the quantum-resistant PIR schemes are based on hard lattice problems such as the (Ring) Learning With Errors (LWE) problem. However, it is important that alternatives to lattice-based PIR schemes are developed to account for possible improvements in lattice problem solvers. One possible alternative are schemes based on hard problems from coding theory. The first code-based PIR scheme has been presented by Holzbaur, Hollanti and Wachter-Zeh in [5] and has been attacked shortly after in [6].

In this paper, we study the code-based PIR scheme presented by Bodur, Martínez-Moro and Ruano in [7]. To prevent the attack in [6], the authors changed the settings to codes over rings in their PIR scheme [7]. We present an attack of this ring-linear cPIR scheme by modifying the approach in [6]: Given the query matrix of the user, the server is able to recover the index of the desired file with high probability in polynomial time.

The PIR Scheme

In a single-server PIR scheme, there is one database storing t files and a user who wants to retrieve the d^{th} file for some $d \in \{1, \dots, t\}$. In order to do so, the user sends a *query* q to the server and the server uses the query and the files in the database to compute the *reply* r. From this reply r, the user can recover the d^{th} file while the server should not learn anything about the index of the requested file.

Let $m = \prod_{i=1}^{\ell} p_i^{e_i}$ where p_i is prime and e_i is a positive integer for all $1 \leq i \leq \ell$. Let $m' = \prod_{i=1}^{\ell} p_i$ and assume that the alphabet of the database is $\mathbb{Z}_{m'}$ (i.e., the ring of integers modulo m'). The database contains t files and we denote by $d \in \{1, \dots, t\}$ the index of the desired file. Each file is stored as a matrix of size $L \times r$ for some positive integers L and r. Due to space limitations, we will only illustrate how the query matrix Q is constructed as this is important for the attack. Q is the expansion in $\mathbb{Z}_m$ of a matrix Q' whose entries are elements in $R := \mathbb{Z}_m/\langle x^n - 1\rangle$, where n is a positive integer such that $\gcd(m, n) = 1$. More precisely, Q' is the concatenation of two matrices of the same size, i.e., $Q' = [\Delta|A]$ where A is a random matrix with entries in $m'R$ and $\Delta = W + E + U$. The matrix Δ is constructed by vertically stacking t matrices $\Delta^i = W^i + E^i + U^i$ on top of each other for $1 \leq i \leq t$, where each of these matrices is of size $r \times s$ for some $s \geq r$.

The rows of W are codewords of a code in R^s, called C_{OUT}, and are generated by computing $W = A \cdot G_{\text{OUT}}$ where G_{OUT} is a generator matrix of C_{OUT}. The matrix E is a random matrix with entries in the non-free part of another code in R, called C_{IN}, meaning that all these entries are multiples of m'. The matrix U only has r non-zero entries, i.e., each row of U^d contains exactly one non-zero entry,

whereas $U^i = 0$ for all $i \neq d$. These non-zero entries are not in C_{IN} and are chosen in such a way that $\text{rowspan}(U) \subseteq C_{\text{OUT}}$.
Finally, Q is obtained by writing the entries of Q' as the corresponding vectors in $\mathbb{Z}_m^n$.

Remark. We want to remark that it is not always possible to uniquely recover the desired file, as there exist several solutions to the resulting system. We also provide an additional condition on the scheme to ensure its completeness.

The proposed Attack
Note that m, m', t, L, r and Q (and therefore also Δ and A) are publicly known. The general idea of the attack is to first remove W in $\Delta = W + E + U$ as $\text{rowspan}(W)$ and $\text{rowspan}(U)$ both lie in the code C_{OUT}. By construction, $\text{rowspan}(E)$ is contained in the code $\Gamma_{\text{IN}} \coloneqq \{(c_1, \ldots, c_s) \mid c_i \in C_{\text{IN}},\, 1 \leq i \leq s\}$ whereas $\text{rowspan}(U)$ is not. Due to this fact, we can then proceed similarly as in [6].

In fact, let $1 \leq i \leq \ell$ and consider the ring $\mathbb{Z}_{p_i^{e_i}}$. From now on, all matrices and computations will be considered over $\mathbb{Z}_{p_i^{e_i}}$. Denote by $A[j]$ the submatrix of A obtained by deleting the rows corresponding to a^j, i.e., the rows $[(j-1)r + 1, jr]$, for $1 \leq j \leq t$. Similarly, we define $\Delta[j], E[j]$ and $U[j]$.
We can consider $A[j]^\top$ as generator matrix of a linear code C and compute a corresponding parity-check matrix $H[j]$ in standard form. Then, $H[j] \cdot A[j] = 0$ and hence,

$$Z[j] \coloneqq H[j] \cdot \Delta[j] = H[j] \cdot (A[j]G_{\text{OUT}} + E[j] + U[j]) = H[j] \cdot (E[j] + U[j]).$$

We can show that, with high probability, these matrices $Z[j]$ are non-zero.

Lemma 1. Let A be a random matrix in $p_i \mathbb{Z}_{p_i^{e_i}}^{rt \times ns}$ and let $A[j]$ and $H[j]$ be constructed as described above, where $1 \leq j \leq t$. If

$$t > \frac{2ns}{r} + 1,$$

then $Z[j] \neq 0$ with high probability.

Note that by definition of U,

$$Z[j] = \begin{cases} H[j] \cdot E[j], & \text{if } j = d \\ H[j] \cdot (E[j] + U[j]), & \text{else.} \end{cases}$$

Therefore, we can show that with high probability, the $\mathbb{Z}_{p_i}$-dimension of $Z[d]$ is lower than the $\mathbb{Z}_{p_i}$-dimension of $Z[j]$ for $j \neq d$. For simplicity, we denote the $\mathbb{Z}_{p_i}$-dimension by $\dim_i(\cdot)$.

Proposition 2. Let K be the rank of the non-free part of Γ_{IN}. If

$$t \geq \frac{K + ns}{r} + 2,$$

then we show that with high probability $\dim_i(Z[j]) > \dim_i(Z[d])$ for all $1 \leq j \leq t$ such that $j \neq d$.

Hence, for every $1 \leq i \leq \ell$, we compute the set $S_i \coloneqq \min\{\dim_i(Z[j]) \mid 1 \leq j \leq t\}$ and can then show that with high probability

$$\bigcap_{i=1}^{\ell} S_i = \{d\}.$$

The cost of the attack is roughly approximated to be in $\mathcal{O}(\ell \cdot t \cdot r^3 (t-1)^3)$. Moreover, the computations can be parallelized since the sets S_i can be computed independently.

We showed that the attack in [6] can be adapted to rings by comparing the $\mathbb{Z}_{p_i}$-dimension of the matrices $Z[j]$ instead of their ranks. This attack is successful with high probability if the number of files t is large enough, i.e., if it is above a certain lower bound. Consequently, replacing fields by rings does not prevent the rank difference attack from [6].

Acknowledgements

This work has been supported by funding from Agentur für Innovation in der Cybersicherheit GmbH.

References

[1] C. Dong and L. Chen, "A fast single server private information retrieval protocol with low communication cost," in *European symposium on research in computer security*, pp. 380–399, Springer, 2014.

[2] E. Kushilevitz and R. Ostrovsky, "Replication is not needed: Single database, computationally-private information retrieval," in *Proceedings 38th annual symposium on foundations of computer science*, pp. 364–373, IEEE, 1997.

[3] H. Lipmaa and K. Pavlyk, "A simpler rate-optimal CPIR protocol," in *International conference on financial cryptography and data security*, pp. 621–638, Springer, 2017.

[4] J. P. Stern, "A new and efficient all-or-nothing disclosure of secrets protocol," in *International conference on the theory and application of cryptology and information security*, pp. 357–371, Springer, 1998.

[5] L. Holzbaur, C. Hollanti, and A. Wachter-Zeh, "Computational code-based single-server private information retrieval," IEEE International Symposium on Information Theory (ISIT), pp. 1065–1070, June 2020.

[6] S. Bordage and J. Lavauzelle, "On the privacy of a code-based single-server computational PIR scheme," vol. 13, no. 4, pp. 519–526, 2021.

[7] Ş. Bodur, E. Martínez-Moro, and D. Ruano, "Single server private information retrieval protocols with codes over rings," *Journal of Algebra and Its Applications*, vol. 24, no. 13n14, p. 2541012, 2025.