УНИВЕРСИТЕТ ИТМО

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
март–апрель 2016 Том 16 № 2 ISSN 2226-1494 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
March–April 2016 Vol. 16 No 2 ISSN 2226-1494 http://ntv.ifmo.ru/en

COMPUTATIONALLY EFFICIENT PRIVATE INFORMATION RETRIEVAL PROTOCOL

A.V. Afanasyeva[a], S.V. Bezzateev[a]

[a] Saint Petersburg State University of Aerospace Instrumentation, Saint Petersburg, 190000, Russian Federation
Corresponding author: bsv@aanet.ru

Article info
Received 23.11.15, accepted 26.01.16
doi: 10.17586/2226-1494-2016-16-2-290-294
Article in English

For citation: Afanasyeva A.V., Bezzateev S.V. Computationally efficient private information retrieval protocol. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2016, vol. 16, no. 2, pp. 290–294, doi:10.17586/2226-1494-2016-16-2-290-294

Abstract

This paper describes a new computationally efficient private information retrieval protocol for one q-ary symbol retrieving. The main advantage of the proposed solution lies in a low computational complexity of information extraction procedure, as well as the constructive simplicity and flexibility in choosing the system parameters. Such results are based on cosets properties. The proposed protocol has communication complexity slightly worse than the best schemes at the moment, which is based on locally decodable codes, but it can be easily built for any parameters of the system, as opposed to codes. In comparison with similar solutions based on polynomials, the proposed method gains in computational complexity, which is important especially for servers which must service multiple requests from multiple users.

Keywords

private information retrieval protocol, polynomial interpolation, Galois group cosets over the finite fields, Lagrange, locally decodable codes

УДК 004.056.5

ВЫЧИСЛИТЕЛЬНО-ЭФФЕКТИВНЫЙ ПРОТОКОЛ КОНФИДЕНЦИАЛЬНОГО ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ

А.В. Афанасьева[a], С.В. Беззатеев[a]

[a] Санкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, 190000, Российская Федерация
Адрес для переписки: bsv@aanet.ru 9214215187@ya.ru

Информация о статье
Поступила в редакцию 23.11.15, принята к печати 26.01.16
doi:10.17586/2226-1494-2016-16-2-
Язык статьи – английский

Ссылка для цитирования: Афанасьева А.В., Беззатеев С.В. Вычислительно-эффективный протокол конфиденциального извлечения информации // Научно-технический вестник информационных технологий, механики и оптики. 2016. Т. 16. № 2. С. 290–294. doi:10.17586/2226-1494-2016-16-2-290-294

Аннотация

Предложен новый вычислительно-эффективный протокол конфиденциального извлечения информации из удаленной базы данных. Основное достоинство предлагаемого решения состоит в низкой вычислительной сложности процедуры извлечения информации, а также в конструктивной простоте и гибкости выбора параметров системы. Результаты получены благодаря использованию свойств орбит действия групп Галуа конечных расширений поля *GF*(*q*). Наилучшим существующим на данный момент кодовым решениям схема уступает по коммуникационной сложности незначительно, но при этом имеет конструктивную процедуру построения для любых допустимых параметров. По сравнению с существующими решениями, основанными на свойствах полиномов, предложенный протокол имеет меньшую вычислительную сложность, что, безусловно, является важным фактором для серверной части, которая должна обслуживать множественные заявки.

Ключевые слова

протокол конфиденциального извлечения информации, интерполяция полиномов, Лагранж, орбиты действия групп Галуа конечных расширений поля, локально-декодируемые коды

Introduction

The private information retrieval (PIR) concept was proposed by Chor, Goldreich, Kushilevitz and Sudan [1]. Authors were the first who has considered the problem of anonymity between data owner and data consumer, from the point of view of the user's security. They formalize the following problem: user would like receive some data from database without revealing its interest to database owner. More formal the problem could be presented in a following way: server holds N-bit string $\mathbf{X}$, a user wishes to retrieve x_i and keeps i private, without requesting all N-bits.

Complexity of PIR schemes includes two components: computation complexity is presented by a server's costs for calculating answer on the user's query by whole database; and communication complexity is presented by network overhead including lengths of queries and answers.

Existing Schemes

All known PIR protocols realization can be divided into two classes: single-server protocols (see [2–10]) and multi-servers ones (see [11–14]).

Single-server solutions have two main advantages: they do not require replication of the original database, and among them there are solutions with the lowest known communication complexity at the moment. However, they also have several disadvantages: for all these systems, security is based on the assumptions of the computational hardness of some problems; in addition, communicatively effective solutions have high computational complexity of both server and client side, which greatly reduces their practical applicability, as they either require the use of expensive hardware or time-consuming.

The multi-server systems can be divided into two classes: the first is based on the arithmetic of polynomials, and the second - on locally decodable codes. Multi-servers PIRs have the following advantages: they can be proved as information-theoretic secure, they have low computational complexity (both on the client and on the server side), that is essential for practical implementation. Minimum known communication complexity provides an approach based on locally decodable codes (LDC), but at the moment there is no constructive algorithm for generating such codes with arbitrary parameters.

By estimating asymptotic behavior of communication complexities it could be concluded that PIRs from matching vectors codes are the most efficient construction, but there is no constructive procedure generating code for any parameters. There are only some examples of codes, and proof of existence, but the only known approach of code's construction is the exhaustive search. The main goal of proposed solution is to minimize calculation complexity without significant lost in communication complexity. The main advantage of proposed approach is constructive procedure for scheme with any parameters. There is description of this procedure in next section by the *Initialization Stage.*

Description of new private information retrieving protocol

The proposed PIR is described according to standard scheme stages.

Stage 1. Initialization

An element $\mathbf{u}_j$ is associated with every j-th position $j \in \{1,2,\dots,n\}$ of data vector $\mathbf{X} = \{x_1, x_2, , x_n\}$ that is $\mathbf{u}_j$ treated as a vector of Hamming weight w and length l:

$$j \to \mathbf{u}_j = (u_j^{(0)}, u_j^{(1)}, \dots, u_j^{(l-1)}), \qquad u_j^{(i)} \in \{0,1\} \subseteq GF(q), wt(\mathbf{u}_j) = w, \binom{l}{w} \geq n.$$

Set $r = w + 1$, where r is the number of servers from which we can obtain responds.

As in previous works [11, 12] each vector $\mathbf{u}_j$ is mapped to monomial $m_j = z_0^{u_j^{(0)}} z_1^{u_j^{(1)}} \cdots z_{l-1}^{u_j^{(l-1)}}$.

Now the database can be described in the following way:

$$\mathbf{X} \to F(z_0, z_1, \dots, z_{l-1}) = \sum_{j=1}^{n} x_j m_j = \sum_{j=1}^{n} x_j \prod_{i=0}^{l-1} z_i^{u_j^{(i)}} . \qquad (1)$$

The last selected characteristic on this stage is field extension $GF(q^m)$. The value of variable m depends on r and should satisfy the inequality:

$$(w+1)m \leq \left|\left\{ \alpha \in GF(q^m) | GF(q)(\alpha) = GF(q^m) \right\}\right|.$$

For each server S_i the element $\alpha_i \in GF(q^m)$ should be choosen in such a way that provide its own coset $O_i = \{\alpha_i^{q^k} | k = 0, \dots, m-1\}$.

Pre-calculate Lagrange coefficients $\left\{\lambda_{01}, \lambda_{01}, \dots, \lambda_{0(mw+1)}\right\}$ for interpolation of polynomial $F(\gamma): \deg(F(\gamma)) \leq wm$ in a point $x = 0$ by standrd Lagrange formulas.

Stage 2. Query generation

A random matrix C of the size $m \times l$ over Z_q is generated to retrieve j-th q-ary block x_j (1). This matrix is common for all servers. The query for each server S_i is formed by using this matrix $\mathbf{C}$. The following steps should be done:

1. To generate basis $\mathbf{B}_i$ for server S_i by using element α_i:

$$\mathbf{B}_i = \left[\alpha_i, \alpha_i^2, \alpha_i^3, \ldots, \alpha_i^m\right].$$

2. To construct matrix $\mathbf{U_j}$ for binary vector $\mathbf{u}_j = \left(u_j^{(0)}, u_j^{(1)}, \ldots, u_j^{(l-1)}\right)$, which is the mapping of requested position j.

$$\mathbf{U_j} = \left[u_j^{(0)}\alpha^0, u_j^{(1)}\alpha^0, \ldots, u_j^{(l-1)}\alpha^0\right].$$

If an element $\alpha^0 \in GF(q^m)$ of Galois field $GF(q^m)$ is presented as a column of size m from $GF(q)$**.** And if we use the basis of $GF(q^m)$ over $GF(q)$ with first element 1.

We receive the following matrix:

$$\mathbf{U_j} = \begin{bmatrix} u_j^{(0)} & u_j^{(1)} & \ldots & u_j^{(l-1)} \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix}.$$

3. Now we can calculate the request matrix by the formula:

$$\mathbf{R_i} = \mathbf{U_j} + \mathbf{B_i C}.$$

The resulting request matrix $\mathbf{R_i}$ can be expressed as a vector of the length l over $GF(q^m)$ in a following way:

$$\mathbf{R_i} = \left[u_j^{(0)}\alpha^0 + \sum_{k=1}^{m} c_{k1}\alpha_i^k, \quad u_j^{(1)}\alpha^0 + \sum_{k=1}^{m} c_{k2}\alpha_i^k, \quad \ldots \quad, u_j^{(l-1)}\alpha^0 + \sum_{k=1}^{m} c_{kl}\alpha_i^k\right],$$

where c_{kj} is an element of k-th row and j-th column of matrix $\mathbf{C}$.

Query is sent to corresponding server.

Stage 3. Computation of server response

Each server S_i calculates the value of the function $F_X(z_0, \ldots, z_{l-1})$ at the point $\mathbf{R_i} = (r_{i1}, \ldots, r_{il})$ and returns to a user the result as an element of $GF(q^m)$.

Stage 4. Bit retrieving

The responding results that are received from the servers are the values of polynomial $F(\gamma)$ points $(\alpha_1, \alpha_2,, \alpha_{w+1})$ are used for the reconstruction of the polynomial $F(\gamma)$: $deg(F(\gamma)) \leq wm$, and unknown values of the polynomial required for the Lagrange interpolation procedure are calculated considering the polynomial properties in $GF(2^m)$ field with the equation $F(x^2) = (F(x))^2$.Thus, for each received value $F(R_h)$ additional $m - 1$ values are calculated at different additional points that are the elements of corresponding coset C^h and the requested bit $x_i = F(0)$ is calculated by the Lagrange interpolation procedure using precalculated Lagrange coefficients $\{\lambda_{01}, \lambda_{02},, \lambda_{0(wm+1)}\}$.

User receives $w + 1$ values of function $F(R_i), i = 1, \ldots, w + 1$ in different points from $w + 1$ Servers. As all elements of request vector $\mathbf{R_i}$ are functions of α_i then user can consider received values not as values of function $F(z_0, z_1, \ldots, z_{l-1})$ of l variables, but as values of function $\hat{F}(x)$ of one variable over the field $GF(q)$ in a point $x = \alpha_i$.

The user can calculate additional values of the function $\hat{F}(x)$ in $(m - 1)(w + 1)$ different points by using cosets $O_i, i = 1, \ldots, w + 1$ properties. User calculates the points by the following formulas:

$$\hat{F}(\alpha_i^q) = (\hat{F}(\alpha_i))^q, \hat{F}(\alpha_i^{q^2}) = (\hat{F}(\alpha_i))^{q^2}, \ldots, \hat{F}(\alpha_i^{q^{m-1}}) = (\hat{F}(\alpha_i))^{q^{m-1}}.$$

As a total resulting $m(w + 1)$ values of the polynomial $\hat{F}(x)$ at different points could be received.

The value $\hat{F}(0)$ could be interpolated by Lagrange polynomial using this values. The value $\hat{F}(0)$ is equal to retrieving block x_j.

Competitive analysis

For competitive comparison of proposed scheme with existing solution all significant parameters are presented in the table. All results for competitive solutions are taken from original papers. The parameters for our scheme can be easily evaluated from the description.

It is possible to conclude from this table, that proposed solution has better computation complexity then Woodruff-Jekhanin scheme and congruent quantity of communication complexity.

The best communication complexity has Matching vectors based scheme. To compare both schemes lets estimate

$$\lim_{N\to\infty}\frac{N^{\frac{1}{r}}}{2^{(\log N)^{1/2}(\log\log N)^{1-1/2}}}$$

for the case $r = 3$. This limit equals to $+\infty$ that shows that in asymptotic the nominator grows faster then denominator. So, the comparison of both schemes shows that in asymptotic communication complexity of our solution grows faster then one of Matching vectors. But non asymptotic comparison of communication overhead shows that LDC based approach is better from DB sizes starting from 2^{50} bits.

Parameters	Woodruff-Jekhanin Scheme [12]	Matching vectors approach [11] $r = 3 \cdot 2^{t-2}$	Our solution
Communication complexity	$O(r^2 \log_2 r N^{1/(2r-1)})$	$2^{(\log N)^{1/t}(\log\log N)^{1-1/t}}$	$O(N^{\frac{1}{r}})$
Storage complexity	N	$2^{2^{(\log N)^{1/t}(\log\log N)^{1-1/t}}}$	N
Computation complexity: – server side	$O(r^2 N^{2r/(2r-1)})$	$O(1)$	$O(rN)$
– client side	$O(r^2 N^{1/(2r-1)})$	$O(r^3)$	$O(r^2)$

Table. Comparison of proposed scheme with existing solution

Security analysis

Since the matrix $\mathbf{C}$ is randomly chosen from a uniform distribution over $F_q^{m\times l}$ and $\mathbf{B_i}$ is a basis of $GF(q^m)$, the matrix $\mathbf{R_h}$ of each server is distributed uniformaly over $F_q^{m\times l}$. By analogy with Shamir secret sharing scheme [15] we can show that each server after receiving $\mathbf{R_h}$ can calculate unique valid $\mathbf{C}$ for all possible $\mathbf{U_j}$. So, even computationally unrestricted adversary can't obtain any information even having knowledge about it.

Conclusion

In this paper the authors proposed the new approach to private information retrieving protocol construction. A new PIR scheme is described with usage of proposed approach. All significant parameters of proposed scheme have been analyzed and compared with existing solutions. As the result of comparison the following conclusions can be done.

- Proposed solution has better or equal storage complexity then all concurrent solution.
- New scheme has better client side computation complexity.
- Proposed PIR algorithm just insignificantly loses in communication complexity.
- New approach allows reaching balance between algorithm parameters.

References

1. Chor B., Kushilevitz E., Goldreich O., Sudan M. Private information retrieval. *Proc. 36th Annual IEEE Symp. Foundation of Computer Science*. Milwaukee, USA, 1995, pp. 41–50.
2. Chor B., Gilboa N. Computationally private information retrieval. *Proc. 29th Annual ACM Symposium on Theory of Computing*. El Paso, USA, 1997, pp. 304–313.
3. Kushilevitz E., Ostrovsky R. Replication is not needed: single database, computationally-private information retrieval. *Proc. 38th IEEE Annual Symposium on Foundations of Computer Science*. Miami Beach, USA, 1997, pp. 364–373.
4. Cachin C., Micaliy S., Stadlerz M. Computationally private information retrieval with polylogarithmic communication. *Lecture Notes in Computer Science*, 1999, vol. 1592, pp. 402–414. doi: 10.1007/3-540-48910-X_28
5. Chang Y.-C. Single database private information retrieval with logarithmic communication. *Lecture Notes in Computer Science*, 2004, vol. 3108, pp. 50–61. doi: 10.1007/978-3-540-27800-9_5
6. Gentry C., Ramzan Z. Single-database private information retrieval with constant communication rate. *Proc. 32th International Colloquium on Automata, Languages and Programming*. Lisbon, Portugal, 2005, pp. 803–815.
7. Melchor C., Gaborit P. A lattice-based computationally-efficient private information retrieval protocol. *IACR Cryptology ePrint Archive*, 2007.

8. Smith S.W., Safford D. Practical server privacy with secure coprocessors. *IBM Systems Journal*, 2001, vol. 40, no. 3, pp. 683–695.
9. Asonov D., Freytag J.C. Almost optimal private information retrieval. *Proc. 2nd Workshop on Privacy Enhancing Technologies*, 2002, vol. 2482, pp. 209–223.
10. Ambainis A. Upper bound on the communication complexity of private information retrieval. *Lecture Notes in Computer Science*, 1997, vol. 1256, pp. 401–407.
11. Yekhanin S. Locally decodable codes. *Lecture Notes in Computer Science*, 2011, vol. 6651, pp. 289–290.
12. Woodruff D., Yekhanin S. A geometric approach to information-theoretic private information retrieval. *SIAM Journal of Computing*, 2007, vol. 37, no. 4, pp. 1046–1056. doi: 10.1137/06065773X
13. Beimel A., Ishai Y., Kushilevitz E. General constructions for information-theoretic private information retrieval. *Journal of Computer and System Sciences*, 2005, vol. 71, no. 2, pp. 213–247. doi: 10.1016/j.jcss.2005.03.002
14. Beimel A., Ishai Y., Kushilevitz E., Raymond J.F. Breaking the O($n^{1/(2k-1)}$) barrier for information-theoretic private information retrieval. *Proc. 43rd Annual IEEE Symposium on Foundations of Computer Science*. Vancouver, Canada, 2002, pp. 261–270.
15. Shamir A. How to share a secret. *Communications of the ACM*, 1979, vol. 22, no. 11, pp. 612–613. doi: 10.1145/359168.359176

Alexandra V. Afanasyeva – senior lecturer, Saint Petersburg State University of Aerospace Instrumentation, Saint Petersburg, 190000, Russian Federation, Alra@vu.spb.ru

Sergey V. Bezzateev – D.Sc., Associate professor, Head of Chair, Saint Petersburg State University of Aerospace Instrumentation, Saint Petersburg, 190000, Russian Federation, bsv@aanet.ru

Афанасьева Александра Валентиновна – старший преподаватель, Санкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, 190000, Российская Федерация, Alra@vu.spb.ru

Беззатеев Сергей Валентинович – доктор технических наук, доцент, заведующий кафедрой, Санкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, 190000, Российская Федерация, bsv@aanet.ru