Порядок коммуникационной сложности PIR-протоколов

Г. А. Майлыбаева

Рассмотрим протокол с $k+1$ участником: пользователем и k несообщающимися серверами ($k \geqslant 1$), причем каждый из серверов хранит один и тот же булев вектор $x = (x_0, \ldots, x_{n-1})$ длины n — базу данных. Пользователь желает узнать значение i-го бита x_i этого вектора так, чтобы номер бита i не стал известен ни одному из серверов. Протокол, который позволяет это делать, называется протоколом доступа к данным без раскрытия запроса или PIR-протоколом и определяется следующим образом.

Для любого натурального n обозначим $E_n = \{0, \ldots, n-1\}$. Пусть k, n, s, $p^0, \ldots, p^{k-1}$ — натуральные числа, $p = p^0 + \ldots + p^{k-1}$. Пусть на множестве $B = \{(i, r),\ i \in E_n,\ r \in E_s\}$ задано вероятностное пространство $\langle B, 2^B, \mathrm{P}\rangle$, где $\mathrm{P}(i, r) = \frac{1}{n \cdot s}$, для любых $i \in E_n$, $r \in E_s$. Тогда (k, n, s, p) *PIR-протоколом* называется набор из $k+2$ функций $I = \langle Q, A^0, \ldots, A^{k-1}, R\rangle$, где $Q, A^0, \ldots, A^{k-1}, R$ некоторые отображения $Q : E_k \times E_n \times E_s \to E_s$, $A^j : E_s \times \{0,1\}^n \to \{0,1\}^{p^j}$, $j \in E_k$, $R : E_n \times E_s \times \{0,1\}^p \to \{0,1\}$, такие, что выполнено 2 условия:

- корректности: для любых $i \in E_n$, $r \in E_s$ выполнено

$$R(i, r, A^0(Q(0, i, r), x), \ldots, A^{k-1}(Q(k-1, i, r), x)) = x_i.$$

- защищенности: для любых $q \in E_s$, $t \in E_k$, $i, j \in E_n$ выполнено

$$\mathrm{P}(Q(t, i, r) = q) = \mathrm{P}(Q(t, j, r) = q).$$

Содержательно протокол $I = \langle Q, A^0, \ldots, A^{k-1}, R\rangle$ состоит из следующих шагов:

- Пользователь U, имея запрос i, вырабатывает случайное число $r \in E_s$, для каждого $j \in E_k$ вычисляет $q^j = Q(j, i, r)$ и посылает q^j j-му серверу S_j.
- Каждый сервер S_j, $j \in E_k$, вычисляет $a^j = (a^j_0, \ldots, a^j_{p^j-1}) = A^j(x, q^j)$ и посылает вектор a^j пользователю.
- U вычисляет $x_i = R(i, r, a^0, \ldots, a^{k-1})$.

Величина $C(I) = k]\log_2 s[+ p$ называется *коммуникационной сложностью протокола* I. $C(I)$ — число бит, переданных в процессе протокола.

Условие корректности гарантирует, что пользователь получит нужный бит базы данных, а условие защищенности — что ни один из серверов по запросу q, который он получил, не сможет понять какой бит интересует пользователя. Предполагается, что всем участникам протокола и пользователю и серверам известны функции запросов, ответов и реконструирующая. Но серверам не известно случайное число r и разумеется не известен номер бита i.

Основной целью исследований в этой области является построение для заданного количества серверов k, длины базы данных n и максимального значения датчика случайных чисел s PIR-протокола с минимальной коммуникационной сложностью.

Степенью существенности булевой функции $f(x_1, \ldots, x_l)$ назовем число переменных, от которых она существенно зависит, и обозначим его через $S(f)$. Степенью существенности булевой вектор-функции $F(x_1, \ldots, x_l) = (f_1(x_1, \ldots, x_l), \ldots, f_t(x_1, \ldots, x_l))$ назовем число $S(F) = \max\limits_{1 \leqslant j \leqslant t} S(f_j)$.

Пусть $A^j(q)(x) = A^j(q, x) = (A^j_0(q, x), \ldots, A^j_{p^j-1}(q, x)), \forall j \in E_k$.

Степенью существенности функции ответов j-го сервера $A^j : E_s \times \{0,1\}^n \to \{0,1\}^{p^j}$, $j \in E_2$, назовем число $S(A^j) = \max\limits_{q \in E_s} S(A^j(q))$.

Нами был найден порядок коммуникационной сложности PIR-протоколов в зависимости от степени существенности функций ответов серверов.

Обозначим через $\mathcal{I}(k, n, s)$ класс всех (k, n, s, p) PIR-протоколов, где $p > 0$. Пусть $\mathcal{A}$ — некоторое множество PIR-протоколов. Тогда обозначим

$$C(k,n,s,\mathcal{A}) = \min\{C(I) : I \in \mathcal{A} \cap \mathcal{I}(k,n,s)\},$$
$$C(k,n,\mathcal{A}) = \min_{s\in\mathbb{N}} C(k,n,s,\mathcal{A}).$$

Для любого натурального d обозначим через $\mathcal{A}_d$ множество всех PIR-протоколов таких, что степень существенности функции ответов каждого сервера не превосходит d.

Теорема 1 (Верхняя оценка). *Для любых натуральных k, n и d таких что $0 < d \leqslant n^{2k-2/2k-1}$, верно*

$$C(k,n,2^{kd^{1/2k-2}},\mathcal{A}_d) \leqslant (k^2+k)d^{1/2k-2} + 2k\frac{n}{d}.$$

Теорема 2 (Нижняя оценка). *Для любых натуральных k, n, s, d верно*

$$C(k,n,s,\mathcal{A}_d) \geqslant k]\log_2 s[+\frac{n}{d}.$$

Следствие 1. *Если натуральные числа k, d, n такие что $0 < d \preccurlyeq n^{2k-2/2k-1}$ при $n \to \infty$, то при $n \to \infty$ верно*

$$C(k,n,\mathcal{A}_d) \asymp \frac{n}{d}.$$

Автор выражает благодарность Э. Э. Гасанову за постановку задачи.

Список литературы

[1] Chor B., Goldreich O., Kushilevitz E., Sudan M. Private information retrieval // Proc. of the 36th Annu. IEEE Symp. on Foundations of Computer Science. 1995. P. 41–51.

[2] Beimel A., Ishai Y., Kushilevitz E., Raymond J.-F. Breaking the $O(n^{1/(2k-1)})$ barrier for information-theoretic private information retrieval // Proc. of the 43st IEEE Sym. on Found. of Comp. Sci. 2002.

[3] Гасанов Э. Э., Майлыбаева Г. А. Доступ к базам данных без раскрытия запроса // Материалы конференции «Математика и безопасные информационные технологии». Москва, 23–24 октября 2003 г. 393–395.

[4] Майлыбаева Г. А. Границы вырожденности протоколов доступа к данным без раскрытия запроса // Дискретная математика. 2006. 18. № 2.