МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
им. М.В.Ломоносова

МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ

На правах рукописи
УДК 519.712.3

Майлыбаева Гульнара Абаевна

Коммуникационная сложность протоколов доступа к данным без раскрытия запроса.

01.01.09 — дискретная математика и математическая кибернетика

автореферат
диссертации на соискание ученой степени
кандидата физико–математических наук

Научный руководитель
доктор физико-математических наук
профессор Э.Э.Гасанов

Москва 2008

Работа выполнена на кафедре математической теории интеллектуальных систем Механико-математического факультета Московского государственного университета имени М.В.Ломоносова.

Научный руководитель:	доктор физико-математических наук, профессор Гасанов Эльяр Эльдарович
Официальные оппоненты:	доктор физико-математических наук, профессор Ложкин Сергей Андреевич
	кандидат физико-математических наук Шакиров Абдыганы Абжамилович
Ведущая организация:	Вычислительный Центр РАН

Защита диссертации состоится 18 апреля 2008 г. в 16 ч. 40 м. на заседании диссертационного совета Д.501.001.84 при Московском государственном университете им. М.В.Ломоносова по адресу: Российская Федерация, 119991, ГСП-1, Москва, Ленинские горы, МГУ, Механико-математический факультет, аудитория 14-08. С диссертацией можно ознакомиться в библиотеке Механико-математического факультета МГУ (Главное здание, 14 этаж).
Автореферат разослан 18 марта 2007 г.

Ученый секретарь диссертационного
совета Д.501.001.84 при МГУ
доктор физико-математических наук,
профессор

А.О.Иванов

Общая характеристика работы

Актуальность темы

В связи с постоянным расширением области применения компьютерных технологий, одним из актуальных направлений дискретной математики и математической кибернетики являются операции хранения и поиска информации. В последние десятилетия активно развивается новое научное направление, связанное с оптимальным хранением и поиском информации, именуемое теорией информационного поиска. Одним из главных носителей этого направления является теория баз данных[1,2]. В данной работе рассматривается одна из составляющих данной теории – проблема защищенности баз данных и поисковых систем, а именно проблема защиты информации в интересах пользователей.

В настоящее время знание некоторых предпочтений пользователей приобретает известное значение и цену. С другой стороны, нет никаких оснований считать, что администратор сервера, на котором хранится база данных, не использует информацию о своем пользователе против его самого.

Протоколы извлечения информации без раскрытия запроса позволяют пользователю получить желаемый бит информации из базы данных таким образом, что администратор базы данных ничего не узнает о номере бита, который запрашивал пользователь. Понятие такого протокола впервые было введено в работе B. Chor, O. Goldreich, E. Kushilevitz и M. Sudan[3] под названием Private Information Retrieval, поэтому мы в дальнейшем будем называть такие протоколы PIR-протоколами.

Существует множество примеров, где использование протоколов, которые скрывают от администраторов базы данных интересы их пользователей (PIR-протоколов), может быть полезно.

Фармацевтические базы данных. Обычно фармацевтические компании специализируются либо в изобретении новых лекарств, либо на

[1]Гасанов Э.Э. О сложности информационного поиска, канд. дисс. Москва, 1985

[2]Гасанов Э. Э., Кудрявцев В.Б. Теория хранения и поиска информации. *Москва, ФИЗМАТЛИТ*, 2002

[3]B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan. Private information retrieval. In Proc. of the 36th Annu. IEEE Symp. on Foundations of Computer Science, pages 41-51, 1995.

сборе информации об определенных компонентах и их свойствах (фармацевтические базы данных). Процесс получения нового лекарства заключает в себе необходимость получения информации из базы данных о его компонентах. Чтобы скрыть планы компании, можно купить целую базу данных. В этом случае PIR-протоколы позволяют избежать огромных затрат.

Учебные примеры. Специальный отдел министерства обороны планирует секретную операцию в регионе X. Чтобы получить карту региона он должен сделать запрос в базу данных карт. Таким образом, может произойти утечка данных о том, что скоро произойдет секретная операция в регионе X. Возможно, конечно, в целях безопасности, купить всю базу данных карт. Опять же, этого возможно избежать при использовании PIR-протоколов.

Таким образом, существуют примеры, когда необходима защита интересов пользователей баз данных. До недавнего времени этот вопрос не учитывался при их построении.

Существует простое решение, когда сервер передает всю базу данных пользователю. Но если считать, что пользователь платит за количество всех принятых и переданных за время протокола бит, то цена такого протокола очень высока. Назовем коммуникационной сложностью PIR-протокола общее количество бит, которыми обмениваются участники за время протокола. Тогда целью построения PIR-протоколов является построение протоколов с минимальной коммуникационной сложностью.

В работе B. Chor, O. Goldreich, E. Kushilevitz и M. Sudan[4] было показано, что если база данных хранится на одном сервере, то минимальная коммуникационная сложность PIR-протокола равна длине базы данных. Чтобы решить эту проблему, было предложено копировать базу данных на k несообщающихся между собой серверах, и проводить протокол таким образом, что каждый отдельный сервер не получает никакой информации об номере искомого бита.

[4]B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan. Private information retrieval. In Proc. of the 36th Annu. IEEE Symp. on Foundations of Computer Science, pages 41-51, 1995.

Рассмотрим протокол с $k+1$ участником: пользователем и k несообщающимися серверами ($k \geq 1$), 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 состоит из следующих шагов.

1) Пользователь имеет номер бита i и вырабатывает случайное число r. По числам i и r пользователь вычисляет с помощью специальной функции, называемой функцией запросов, k чисел q^j и посылает j-му серверу запрос q^j.

2) Каждый из k серверов по полученному запросу q^j и базе данных x с помощью специальной функции ответов вычисляет вектор a^j и посылает его пользователю.

3) Пользователь по числам i, r и k ответам серверов a^j вычисляет с помощью реконструирующей функции нужный бит x_i.

Первое требование к протоколу состоит в том, что ни один из серверов по своему запросу q^j не может понять, с помощью какого бита i этот запрос был порожден. Это требование называется требованием защищенности. Второе требование к протоколу, называемое требованием корректности, заключается в том, что пользователь по ответам серверов правильно восстанавливает бит x_i. 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. Целью построения PIR-протоколов является построение протокола с минимальной коммуникационной сложностью для заданных n и k.

Приведем формальное определение PIR-протокола. Для любого натурального n обозначим $E_n = \{0, \ldots, n-1\}$. Пусть k, n, s, m, $p^0, \ldots, p^{k-1}$ — 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 $< B, 2^B, \mathbf{P} >$, где $\mathbf{P}(i, r) = \frac{1}{n \cdot s}$, для любых $i \in E_n$, $r \in E_s$. Тогда (k, n, s, m, 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, $Q : E_k \times E_n \times E_s \to E_m$, $A^j : E_m \times \{0,1\}^n \to \{0,1\}^{p^j}$, $j \in E_k$, $R : E_n \times E_s \times \{0,1\}^p \to \{0,1\}$, такие,

что выполнены 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 $q \in E_m$, $t \in E_k$, $i, j \in E_n$ выполнено

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

Наиболее известные результаты по этой тематике были получены в работах: С.Еханина[5], А.Разборова и С.Еханина[6], Beimel, Y. Ishai, E. Kushilevitz и J. F. Raymond[7], в O.Goldreich, H.karloff, L.Schulman и L.Trevisan [8] и в работе I.Kerendis и R. de Wolf[9].

Цель работы. В работе рассматривается задача построения PIR-протоколов. Целью работы является исследование коммуникационной сложности PIR-протоколов при различных ограничениях на параметры протокола и на множество функций, разрешенных к использованию в протоколе.

Научная новизна. Исследования, проведенные в данной работе, направлены на изучение коммуникационной сложности PIR-протоколов. В данной работе впервые проведено глубокое исследование задачи построения PIR-протоколов при различных ограничениях на параметры протокола и разрешенные к использованию функции.

Известно, что всегда существует простейший PIR-протокол, коммуникационная сложность которого равна длине базы данных n. В

[5]S. Yekhanin. New Locally Decodable Codes and Private Information Retrieval Schemes, Electronic Colloquium on Computational Complexity (ECCC), TR06-127.

[6]Alexander A. Razborov, Sergey Yekhanin. An Omega($n^{1/3}$) Lower Bound for Bilinear Group Based Private Information Retrieval. FOCS 2006: 739-748.

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

[8]O.Goldreich, H.karloff, L.Schulman, and L.Trevisan. Lower bounds for linear locally decodable codes and private information retrieval systems. In Proc. of the 17th IEEE Conf. on Complexity Theory. IEEE Computer Society Press, 2002.

[9]I.Kerendis and R. de Wolf. Exponential lower bound for 2-query locally decodable codes. In Proc. of the 35th ACM Sym. on Theory of Computing, pages 106-115, 2003.

этом протоколе каждый сервер выдает пользователю свою часть базы данных, а пользователь, собрав всю базу данных извлекает нужный бит. PIR-протокол, у которого коммуникационная сложность больше либо равна длине базы данных, будем считать вырожденным.

В работе впервые была предложена и разрешена проблема принадлежности PIR-протокола к классу невырожденных PIR-протоколов по основным параметрам.

Впервые была получена нижняя оценка коммуникационной сложности для класса PIR-протоколов с более чем 2-мя серверами. Также впервые была получена нетривиальная точная оценка коммуникационной сложности PIR-протоколов.

Описанная в данной работе оценка коммуникационной сложности получена для более широкого класса PIR-протоколов, чем в известных работах по этой тематике. Во-первых, в отличие от полученных ранее результатов, мы не предполагали, что длины ответов серверов должны быть равны между собой, во-вторых, мы не налагаем никаких ограничений на количество бит, которые пользователь использует из ответов серверов. В-третьих, получена нижняя оценка, которая не налагает ограничений на линейность функций используемых в протоколе. И наконец, полученная нижняя оценка коммуникационной сложности по порядку совпадает с коммуникационной сложностью наилучшего известного на сегодняшний момент PIR-протокола для $k = 2$ серверов. Также заметим, что для доказательства известных нижних оценок, в работе O.Goldreich, H.karloff, L.Schulman и L.Trevisan [10] авторы использовали сведение PIR-протоколов к LDC-протоколам, а в работе I.Kerendis и R. de Wolf[11] авторы использовали сведение PIR-протоколов к квантовым PIR-протоколам. Полученная нами нижняя оценка коммуникационной сложности PIR-протоколов доказана напрямую для PIR-протоколов.

Степенью существенности булевой функции $f(x_1, \ldots, x_l)$ назовем число

[10]O.Goldreich, H.karloff, L.Schulman, and L.Trevisan. Lower bounds for linear locally decodable codes and private information retrieval systems. In Proc. of the 17th IEEE Conf. on Complexity Theory. IEEE Computer Society Press, 2002.

[11]I.Kerendis and R. de Wolf. Exponential lower bound for 2-query locally decodable codes. In Proc. of the 35th ACM Sym. on Theory of Computing, pages 106-115, 2003.

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

$$S(F) = \max_{1 \le j \le t} S(f_j).$$

Пусть $A^j(q, x) = (A^j_0(q, x), \ldots, A^j_{p^j-1}(q, x))$. Для функций $A^j(q, x), A^j_l(q, x), l \in E_{p^j}, j \in E_k, q \in E_s$ также будем использовать следующую запись: $A^j(q, x) = A^j(q)(x), A^j_l(q, x) = A^j_l(q)(x)$, где $A^j(q)\vdots\{0,1\}^n \to \{0,1\}^{p^j}$ — булева вектор-функция n переменных, $A^j_l(q)\vdots\{0,1\}^n \to \{0,1\}$ — булева функция n переменных.

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

$$S(A^j) = \max_{q \in E_s} S(A^j(q)).$$

Для любых $i \in E_n, r \in E_s$ определим следующую булеву функцию от p переменных: $R_{i,r} : \{0,1\}^p \to \{0,1\}$, где $R_{i,r}(a^0, \ldots, a^{k-1}) = R(i, r, a^0, \ldots, a^{k-1})$.

Степенью существенности реконструирующей функции назовем число

$$S(R) = \max_{i \in E_n, r \in E_s} S(R_{i,r}).$$

В работе получены следующие основные результаты.

1. Найдены границы вырожденности PIR-протоколов по основным параметрам: по количеству серверов; по мощности множества значений датчика случайных чисел; по мощности множества значений и степени существенности функции запросов; по степени существенности реконструирующей функции. В результате получен критерий принадлежности PIR-протокола к классу невырожденных PIR-протоколов.

2. В классе PIR-протоколов с 2-мя серверами, степень существенности функции ответов которых не превышает 2, и длина датчика случайных чисел достаточно мала, построен PIR-протокол и показано, что в данном классе PIR-протоколов его коммуникационная сложность асимптотически минимальна. Для более узкого подкласса этого класса, когда длина базы данных кратна квадрату длины датчика случайных чисел, получена точная оценка коммуникационной сложности.

   Для любого натурального d, в классе PIR-протоколов с $k \geq 2$ серверами, степень существенности функции ответов которых не превышает d, и длина датчика случайных чисел больше длины базы данных, получена нижняя оценка коммуникационной сложности. На основе нижней оценки получено точное по порядку значение коммуникационной сложности для класса PIR-протоколов со степенью существенности функции ответов d и достаточно большой длины датчика случайных чисел. А именно, построен PIR-протокол и показано, что в данном классе PIR-протоколов его коммуникационная сложность по порядку минимальна.

3. Поскольку во всех известных PIR-протоколах, в результате каждого запроса пользователь, помимо запрашиваемого бита, получает много дополнительной информации о других битах базы данных, он может получить всю базу данных за количество запросов, меньшее ее длины. Степенью раскрытия PIR-протоколом базы данных назовем число шагов, за которое пользователь может вычислить всю базу данных. В работе были исследованы основные известные PIR-протоколы, а также построенные нами PIR-протоколы на их степень раскрытия.

Основные методы исследования. В работе используются методы дискретного анализа, комбинаторики, линейной алгебры и теории сложности управляющих систем.

Апробация работы. Результаты настоящей работы докладывались на семинарах механико-математического факультета МГУ им. М.В.Ломоносова: на семинаре "Теория автоматов" под руководством академика, проф., д.ф-м.н. В.Б. Кудрявцева (2005-2007 гг.), на семинаре "Вопросы сложности

алгоритмов поиска" под руководством проф., д.ф-м.н. Э.Э.Гасанова (2004-2007 гг.), на семинаре факультета вычислительной математики и кибернетики МГУ им. М.В.Ломоносова "Дискретная математика и математическая кибернетика"под руководством проф. В.Б. Алексеева, проф. А.А. Сапоженко, проф. С.А. Ложкина (2007 г.), на конференции "Математика и безопасные информационные технологии" (Москва, 2003 г.), на XIV Международной конференции по проблемам теоретической кибернетики (Пенза, 2005 г.), на первой, второй и третьей научной конференциях аспирантов кафедры МаТИС механико-математического факультета МГУ (Москва, 2005-2007 гг.), на Ломоносовских чтениях МГУ (Москва, 2005-2007 гг.), на международном математическом конгрессе "International Congress of Mathematics" (Мадрид, 2006 г.), на IX международной конференции "Интеллектуальные системы и компьютерные науки" (Москва, 2006 г.) .

Публикации. Основные результаты диссертации опубликованы в 7 работах автора, список которых приводится в конце автореферата [1-7].

Структура и объем работы. Диссертация состоит из введения, трех глав и приложения. Объем диссертации 109 стр. Список литературы содержит 43 наименования.

Краткое содержание работы

Во Введении приводится постановка исследуемой задачи, излагается краткий исторический обзор исследований по теме диссертации, вводится строгое определение PIR-протокола, формулируются основные результаты работы. В частности, во введении приводятся описания всех наиболее известных на сегодняшний момент PIR-протоколов и основные методы их построения. В русскоязычной литературе такой обзор не публиковался. Также во введении описан метод практического применения PIR-протоколов.

В главе 1 исследуются границы вырожденности PIR-протоколов по основным параметрам. В частности, в разделе 1.1 приводится простейший вырожденный PIR-протокол, в разделе 1.2 приводится простейший невырожденный PIR-протокол. В разделах 1.3 - 1.8 приводятся доказательства утверждений о границах вырожденности PIR-протокола по основным параметрам: числу серверов, по длине датчика случайных чисел, по

длине базы данных, по степени существенности функции ответов, по степени существенности реконструирующей функции.

Из утверждений 1.1 - 1.8 следует теорема о границах вырожденности PIR-протокола по основным параметрам.

Теорема 1. *Для любого натурального $n \geq 12$ невырожденный (k, n, s, m, p) PIR-протокол $I = \langle Q, A^0, \ldots, A^{k-1}, R\rangle$ существует тогда и только тогда, когда одновременно выполнены следующие условия:*

1. *количество серверов $k \geq 2$,*
2. *длина датчика случайных чисел $s \geq 2$,*
3. *мощность множества запросов $m \geq 2$,*
4. *существует такое $j \in E_k$, что выполнено $S(A^j) \geq 2$,*
5. *$S(R) \geq 2$.*

PIR-протоколы, для которых выполнено $m = s$ будем обозначать четверкой параметров (k, n, s, p). 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

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

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

В главе 2 исследуется коммуникационная сложность PIR-протоколов. Раздел 2.1 посвящен исследованию коммуникационной сложности PIR-протоколов в классе $\mathcal{A}_2$ — классе PIR-протоколов, степень существенности функции ответов которых не превосходит 2. Раздел 2.1.1 посвящен доказательству леммы о верхней оценке в данном классе, а именно в данном разделе построен PIR-протокол с заданными параметрами.

Лемма 1. *Для любых натуральных n, s таких что $s < \sqrt{n}$ верно*

$$C(2, n, s, \mathcal{A}_2) \leq 2]\log_2 s[+\frac{s+1}{2s}n + n \mod s^2.$$

Назовем булеву функцией $f(x_1, \ldots, x_l)$ линейной, если для любых $x_t^1, x_t^2 \in \{0,1\}, 1 \le t \le l$ верно

$$\begin{aligned} f(x_1, \ldots, x_{t-1}, x_t^1 + x_t^2, x_{t+1}, \ldots, x_l) &= f(x_1, \ldots, x_{t-1}, x_t^1, x_{t+1}, \ldots, x_l) + \\ &+ f(x_1, \ldots, x_{t-1}, x_t^2, x_{t+1}, \ldots, x_l). \end{aligned}$$

Линейным (k, n, s, p) PIR-протоколом назовем такой PIR-протокол, у которого для любых $j \in E_k, l \in p^j, r, q \in E_s, i \in E_n$ функции $A_l^j(q)$ и $R_{i,r}$ являются линейными функциями. Положим $\mathcal{D}_2$ — множество всех линейных PIR-протоколов из класса $\mathcal{A}_2$. Проще говоря, для любого PIR-протокола из $\mathcal{D}_2$ верно: каждый бит ответа каждого сервера является суммой некоторых бит базы данных, и для того, чтобы получить значение искомого бита пользователь складывает некоторые биты ответов серверов. В разделе 2.1.2 доказывается, что в классе $\mathcal{A}_2$ для построения PIR-протоколов можно использовать только линейные функции.

Теорема 2. *Для любого $(2, n, s, p)$ PIR-протокола из класса $\mathcal{A}_2$, существует $(2, n, s, p)$ PIR-протокол из класса линейных протоколов $\mathcal{D}_2$ с такой же коммуникационной сложностью.*

В разделах 2.1.3 и 2.1.4 приведены примеры PIR-протоколов с мощностью датчика случайных чисел равного 2 и 3 соответственно.

В разделе 2.1.5 приводится доказательство леммы о нижней оценке коммуникационной сложности PIR-протоколов в классе $\mathcal{A}_2$.

Лемма 2. *Для любых натуральных n, s таких что $s < n$ верно*

$$C(2, n, s, \mathcal{A}_2) \ge 2] \log_2 s[+ \frac{s+1}{2s} n.$$

Будем писать $\alpha(n) = \bar{o}(1)$, если $\lim\limits_{n \to \infty} \alpha(n) = 0$; $A(n) = \bar{o} \cdot (B(n))$, если $A(n) = B(n) \cdot \bar{o}(1)$. Скажем, что $A(n)$ асимптотически не превосходит $B(n)$ при $n \to \infty$ и обозначим $A \lesssim B$, если существует $\alpha(n) = \bar{o}(1)$ такое, что начиная с некоторого номера n_0, $A(n) \le (1 + \alpha(n)) \cdot B(n)$. Если $A(n) \lesssim B(n)$ и $A(n) \gtrsim B(n)$, то будем говорить что A и B асимптотически равны при $n \to \infty$ и обозначать $A \sim B$.

Из леммы 1 и 2 следует теорема об асимптотически точной оценке коммуникационной сложности PIR-протоколов в классе $\mathcal{A}_2$.

Теорема 3. *Если* $s = o(\sqrt{n})$ *при* $n \to \infty$, *то при* $n \to \infty$ *верно*

$$C(2, n, s, \mathcal{A}_2) \sim \frac{s+1}{2s} n,$$

и при n *кратном* s^2 *верно*

$$C(2, n, s, \mathcal{A}_2) = 2]\log_2 s[+\frac{s+1}{2s} n.$$

Раздел 2.2 посвящен исследованию коммуникационной сложности PIR-протоколов в классе $\mathcal{A}_d$ — классе PIR-протоколов, степень существенности функции ответов которых не превосходит d. В разделе 2.2.1 приводится доказательство теоремы о верхней оценке коммуникационной сложности PIR-протоколов для k серверов, степень существенности функций ответов серверов которых не превосходит d, где $0 < d \leq n^{2k-2/2k-1}$.

Лемма 3 (Верхняя оценка). *Для любых натуральных* k, n *и* d *таких что* $0 < d \leq n^{2k-2/2k-1}$, *верно*

$$C(k, n, 2^{kd^{1/2k-2}}, \mathcal{A}_d) \leq (k^2 + k)d^{1/2k-2} + 4k\frac{n}{d}.$$

В разделе 2.2.2 приводится пример PIR-протокола при $d = \frac{n}{3}, s = 2^{\frac{n}{3}}$.

В разделе 2.2.3 приводится доказательство теоремы о нижней оценке коммуникационной сложности PIR-протоколов для k серверов, степень существенности функций ответов серверов которых не превосходит d, где d – произвольное натуральное число.

Теорема 4 (Теорема о нижней оценке). *Для любых натуральных* k, n, s, d *верно*

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

Из леммы 3 и теоремы 4 следует следующая теорема о порядке коммуникационной сложности PIR-протоколов в классе $\mathcal{A}_d$.

Будем писать $A \preccurlyeq B$, если существует такая положительная константа c, что $A(n) \leq c \cdot B(n)$, начиная с некоторого номера n_0. Если $A \preccurlyeq B$ и $A \succcurlyeq B$, то будем говорить, что A и B равны по порядку при $n \to \infty$ и обозначать $A \asymp B$.

Теорема 5. *Если натуральные числа* k, d, n *такие что* $d \preccurlyeq n^{2k-2/2k-1}$ *при* $n \to \infty$, *то при* $n \to \infty$ *верно*

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

В главе 3 исследуется степень раскрытия PIR-протоколов. В разделе 3.1 приводится описание понятия степени раскрытия PIR-протокола. В результате одного запроса к серверам пользователь получает не только искомый бит, но и некоторую информацию об остальных битах базы данных. Степенью раскрытия назовем наименьшее количество запросов, которое должен сделать пользователь, чтобы в результате из полученных ответов серверов он мог восстановить все биты базы данных.

Определение 1. *Для любого* (k, n, s, p) *PIR-протокола* $I = \langle Q, A^0, \ldots, A^{k-1}, R\rangle$, *будем говорить, что множество пар* $\{(i_0, r_0), \ldots, (i_{t-1}, r_{t-1})\}, t \in E_n$ *и ответы* $A^j(Q(j, i_m, r_m), x) = (a^j_{0,m}, \ldots, a^j_{p^j-1,m}), j \in E_k, m \in E_t$ *раскрывают базу данных* $x = (x_0, \ldots, x_{n-1}) \in E_2^n$, *если система уравнений* $A^j_l(Q(j, i_m, r_m), x) = a^j_{l,m}, j \in E_k, l \in E_{p^j}, m \in E_t$ *имеет единственное решение относительно переменных* $x_i, i \in E_n$.

Определение 2. *Степенью раскрытия базы данных* $x = (x_0, \ldots, x_{n-1}) \in E_2^n$ *помощью* (k, n, s, p) *PIR-протокола* $I = \langle Q, A^0, \ldots, A^{k-1}, R\rangle$ *для заданной последовательности случайных чисел* $\tilde{r} = (r_0, \ldots, r_{n-1}) \in E_s^n$ *назовем минимальное число* $t = t(n, \tilde{r})$, *для которого существуют такая перестановка* $\pi : E_n \to E_n$, *что множество пар* $\{(\pi(0), r_0), \ldots, (\pi(t-2), r_{t-2})\}$ *и ответы*

$$A^0(Q(0, \pi(0), r_0), x), \ldots, A^{k-1}(Q(k-1, \pi(0), r_0), x), \ldots$$
$$\ldots, A^0(Q(0, \pi(t-2), r_{t-2}), x), \ldots, A^{k-1}(Q(k-1, \pi(t-2), r_{t-2}), x)$$

не раскрывают базу данных, а множество пар $\{(\pi(0), r_0), \ldots, (\pi(t-1), r_{t-1})\}$ *и ответы*

$$A^0(Q(0, \pi(0), r_0), x), \ldots, A^{k-1}(Q(k-1, \pi(0), r_0), x), \ldots$$
$$\ldots, A^0(Q(0, \pi(t-1), r_{t-1}), x), \ldots, A^{k-1}(Q(k-1, \pi(t-1), r_{t-1}), x)$$

раскрывают базу данных x.

Обозначим

$$\bar{t}(n) = \max_{\tilde{r} \in E_s^n} t(n, \tilde{r}),$$

$$\underline{t}(n) = \min_{\tilde{r} \in E_s^n} t(n, \tilde{r}).$$

В разделе 3.2 рассматривается протокол $I^{1/3}$ для 2-х серверов, описанный в работе B. Chor, O. Goldreich, E. Kushilevitz и M. Sudan[12]. Приводится доказательство следующей теоремы.

Теорема 6. *Для любого натурального n, такого что $\frac{1}{2}n^{1/3}$ – целое, для $(2, n, s, p)$ PIR-протокола $I^{1/3} = \langle Q, A^0, A^1, R\rangle$ верно*

$$\frac{1}{4}n^{2/3} \leq \underline{t}(E_n) \leq \bar{t}(E_n) \leq \frac{1}{3}n^{2/3} - 2.$$

В разделе 3.3 рассматривается протокол I^{pol} для k серверов, описанный в работе A.Beimel, Y.Ishai и E.Kushilevitz[13]. Приводится доказательство следующей теоремы.

Теорема 7. *Для $(2, n, s, p)$ PIR-протокола $I^{pol} = \langle Q, A^0, A^1, R\rangle$ верно*

$$]\frac{n}{2(m+1)}[\leq \underline{t}(n) \leq \bar{t}(n) \leq \min\{1,]\frac{n}{m+1} - 1[\},$$

где m такое что $C_m^3 \geq n$.

В разделе 3.4 рассматривается протокол I^2 для 2-х серверов, описанный в работе Майлыбаевой Г.А.[14] Приводится доказательство следующей теоремы.

Теорема 8. *Для $(2, n, s, p)$ PIR-протокола $I^2 = \langle Q, A^0, A^1, R\rangle$ верно*

$$\underline{t}(n) = \bar{t}(n) = s.$$

[12]B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan. Private information retrieval. In Proc. of the 36th Annu. IEEE Symp. on Foundations of Computer Science, pages 41-51, 1995.

[13]A.Beimel, Y.Ishai and E.Kushilevitz and E.Kushilevitz. General constructions for information-theoretic private information retrieval, 2003.

[14]Майлыбаева Г.А. Точное значение коммуникационной сложности для одного класса PIR-протоколов. Интеллектуальные системы, (2007) 11, №1-4, 167–200.

В разделе 3.5 рассматривается протокол I^d для 2-х серверов, с мощностью датчика случайных чисел равной s, принадлежащий классу $\mathcal{A}_{log_2 s}$, описанный в в работе Майлыбаевой Г.А.[15] Приводится доказательство следующей теоремы.

Теорема 9. *Для* $(2, n, s, p)$ *PIR-протокола* $I^d = \langle Q, A^0, A^1, R\rangle$ *верно*

$$\frac{\log_2 s}{2} = \underline{t}(n) \leq \bar{t}(n) = \log_2 s.$$

В разделе 3.5 рассматривается протокол I^{d1} для 2-х серверов из класса $\mathcal{A}_d$, где $0 < d < n^{2/3}$ – произвольное натуральное число, описанный в работе Майлыбаевой Г.А. [15] Приводится доказательство следующей теоремы.

Теорема 10. *Для* $(2, n, s, p)$ *PIR-протокола* $I^{d1} = \langle Q, A^0, A^1, R\rangle$ *верно*

$$\frac{1}{4}d = \underline{t}(n) \leq \bar{t}(n) = \frac{1}{3}d - 2.$$

Благодарности.

Я благодарю научного руководителя доктора физико-математических наук, профессора Гасанова Эльяра Эльдаровича за постановку задачи, постоянное внимание и помощь в работе, а также академика, доктора физико-математических наук Кудрявцева Валерия Борисовича за многочисленные полезные советы на всех этапах подготовки диссертации. Я выражаю благодарность коллективу кафедры МаТИС за теплую творческую атмосферу.

Работы автора по теме диссертации

1. Майлыбаева Г.А. Границы вырожденности протоколов доступа к данным без раскрытия запроса. Дискретная математика (2006) 18, N 2, 98-110.

2. Maylybaeva G.A. Degeneracy bounds for private information retrieval protocols. Discrete Mathematics and Applications, Volume 16, Number 3, 2006, pp. 245-257.

[15]Майлыбаева Г.А. Порядок коммуникационной сложности для одного класса PIR-протоколов. Дискретная математика.

3. Майлыбаева Г.А. Оценки коммуникационной сложности линейных PIR-протоколов. Интеллектуальные системы, (2005) 9, №1-4, 561–562.

4. Gulnara A. Maylybaeva, Communication complexity for a special class of private information retrieval protocols, In proc. of ICM2006, August (2006), pp. 499.

5. Майлыбаева Г.А. Коммуникационная сложность протоколов доступа к данным без раскрытия запросов. Материалы IX Международной конференции "Интеллектуальные системы и компьютерные науки"(Москва, 23-27 октября 2006 г.), том 1, часть 1, с. 181-183.

6. Майлыбаева Г.А. Точное значение коммуникационной сложности для одного класса PIR-протоколов. Интеллектуальные системы, (2007) 11, №1-4, 167–200.

7. Майлыбаева Г.А. Порядок коммуникационной сложности PIR-протоколов. Интеллектуальные системы, (2007) 11, №1-4, 729–732.