УДК 621.391

КОДЫ ГОППЫ В ПРОТОКОЛАХ АНОНИМНОГО ЗАПРОСА К ДАННЫМ

С. В. Беззатеев,
канд. техн. наук, доцент
Санкт-Петербургский государственный университет аэрокосмического приборостроения

Предложен протокол анонимных запросов к данным, использующий семейства вложенных кодов Гоппы.

Ключевые слова *— распределенные базы данных, анонимность запросов, коды Гоппы.*

Введение

В настоящее время информация о предпочтениях и интересах частных лиц представляет определенный интерес и в некоторых случаях имеет вполне значимую цену. При этом, к сожалению, нельзя гарантировать, что владельцы информационных ресурсов будут соблюдать конфиденциальность и не будут собирать данные о предпочтениях клиентов с целью перепродать их третьим лицам. Протоколы извлечения информации без раскрытия запроса позволяют пользователю получить желаемую информацию из базы данных (БД) таким образом, что владелец БД ничего не узнает о номере бита, который запрашивал пользователь.

Понятие такого протокола впервые было введено в работе [1] под названием *Private Information Retrieval Protocol*, поэтому мы в дальнейшем будем называть такие протоколы «протоколами анонимного запроса к данным». Существует множество примеров, когда использование протоколов, которые скрывают от владельца БД интересы клиентов, пользующихся такими базами, может быть полезно. Приведем здесь лишь некоторые из таких систем.

Фармацевтические БД. Обычно фармацевтические компании интересуются либо вопросами изобретения и соответственно патентования новых лекарств, либо получением максимального объема информации об определенных компонентах лекарственных препаратов и их свойствах (фармацевтические БД). В процессе создания нового лекарства производителю требуется получить максимальный объем информации о свойствах его компонентов. По запросам сотрудников компании-производителя можно получить представление о том, разработкой какого типа лекарств занимается компания в данный момент. Чтобы скрыть планы компании, можно купить все необходимые для работы БД, однако это весьма дорого и, кроме того, требует постоянных обновлений. В этом случае использование протоколов анонимного запроса к данным позволяет избежать таких затрат.

Собственные распределенные БД. В настоящее время все более востребованными становятся распределенные хранилища данных, предоставляемые в пользование различным организациям. Такие услуги существенно снижают расходы компаний на хранение информации и поддержание работоспособности систем управления БД. Однако в таком случае появляется проблема конфиденциальности хранящейся и обрабатываемой таким образом информации. В этом случае защищаемой информацией является не только содержание хранимых и обрабатываемых данных, но и то, к каким данным обращаются пользователи наиболее часто или в какие моменты времени. Помимо этого, третьи лица могут интересоваться закономерностями в обращениях к определенным данным и определенным, наиболее частым связям запросов между собой. Использование для таких систем протоколов анонимного запроса позволяет обеспечить конфиденциальность не только самих данных в неконтролируемых хранилищах, но и скрыть информацию о запросах к таким БД.

Использование семейства вложенных кодов Гоппы для обеспечения анонимности запросов

Одним из способов построения протоколов анонимных запросов является использование помехоустойчивых кодов. В работах [2–4] для этих целей предлагается использовать специальный класс кодов — так называемые частично декодируемые

коды, т. е. коды, в которых можно восстанавливать определенную часть информационных символов. Здесь будет рассмотрен вариант построения протокола анонимных запросов, использующий семейства вложенных кодов Гоппы [5].

Данные, находящиеся в информационной базе, предварительно будут закодированы заранее выбранным множеством вложенных кодов Гоппы. Для каждого i-го поля записи выбирается свой код Гоппы с многочленом $G_i(x)$ и множеством нумераторов позиций L_i в качестве «базового». Информация из i-го поля кодируется с использованием надкода $(L_i, g_i(x))$, где $G_i(x) \equiv 0 \bmod g_i(x)$ и минимальные расстояния этих кодов соотносятся следующим образом:

$$d_i \le (d(\Gamma_i) - 1)/2,$$

где $d(\Gamma_i)$ — минимальное расстояние базового кода Гоппы $(L_i, G_i(x))$.

Таким образом, слова минимального веса надкода $(L_i, g_i(x))$ можно интерпретировать как векторы ошибок, которые могут быть исправлены «базовым $(L_i, G_i(x))$-кодом» Гоппы, выбранным для данного поля записи.

Принцип кодирования с использованием «базового» $(L_i, G_i(x))$-кода и его надкода $(L_i, g_i(x))$ выглядит в общем случае следующим образом. Информация, представленная кодовым словом $\mathbf{a}_1$ веса d_i надкода $(L_i, g_i(x))$, маскируется произвольным кодовым словом $\mathbf{b}_i$ «базового» кода Гоппы $(L_i, G_i(x))$:

$$\mathbf{e}_1 = \mathbf{a}_1 + \mathbf{b}_i.$$

В результате запись, состоящая из n различных полей, будет иметь вид $[\mathbf{e}_1\ \mathbf{e}_2\ \dots\ \mathbf{e}_n]$.

Очевидно, что в каждом поле информация представлена некоторым произвольным кодовым словом соответствующего надкода $(L_i, g_i(x))$. Запрос — *request* — на получение значения определенного k-го поля записи, обеспечивающий при этом анонимность запрашиваемого поля, будет выглядеть следующим образом:

$$request = [\mathbf{h}^*_1\ \mathbf{h}^*_2\ \dots\ \mathbf{H}^*_k\ \dots\ \mathbf{h}^*_n],$$

где $\mathbf{h}^*_i = \mathbf{A}_i \cdot \mathbf{h}_i \cdot \mathbf{P}_i$, здесь $\mathbf{A}_i$ — случайная матрица размером $r_G \times r_{gi}$, r_G — избыточность «базового» (L_i, G_i)-кода, r_{gi} — избыточность надкода $(L_i, g_i(x))$; $\mathbf{h}_i$ — проверочная матрица надкода размером $r_{gi} \times n_i$; $\mathbf{P}_i$ — перестановочная матрица размером $n_i \times n_i$, n_i — длина «базового» кода $(L_i, G_i(x))$;

$$\mathbf{H}^*_k = \mathbf{A}_k \cdot \mathbf{H}_k \cdot \mathbf{P}_k,$$

здесь $\mathbf{A}_k$ — случайная не особенная матрица размером $r_{Gk} \times r_{Gk}$; $\mathbf{H}_k$ — проверочная матрица «базового» кода размером $r_{Gk} \times n_k$; $\mathbf{P}_k$ — перестановочная матрица размером $n_k \times n_k$, n_k — длина надкода $(L_k, g_k(x))$, совпадающая с длиной «базового» $(L_k, G_k(x))$-кода.

Нетрудно доказать, что в результате выполнения запроса к информации, содержащейся в k-м поле записи при использовании выписанного вектора запроса *request*, получим

$$[\mathbf{e}_1\ \mathbf{e}_2\ \dots\ \mathbf{e}_n] \times [\mathbf{h}^*_1\ \mathbf{h}^*_2\ \dots\ \mathbf{H}^*_k\ \dots\ \mathbf{h}^*_n]^T = \mathbf{b}_k \cdot \mathbf{H}_k{}^{\mathrm{T}} \cdot \mathbf{A}_k{}^T,$$

где $\mathbf{b}_k$ — кодовое слово надкода $(L_k, g_k(x))$, вес которого не превышает корректирующей способности «базового» $(L_k, G_k(x))$-кода Гоппы.

Зная матрицу $\mathbf{A}_k$, легко получить значения синдромных компонент $\mathbf{R}_k$ для «базового» $(L_k, G_k(x))$-кода:

$$\mathbf{b}_k \cdot \mathbf{H}_k{}^T \cdot \mathbf{A}_k{}^T \cdot (\mathbf{A}_k{}^{-1})^T = \mathbf{b}_k \cdot \mathbf{H}_k{}^T = \mathbf{R}_k.$$

Далее, зная множество нумераторов позиций L_k и многочлен Гоппы $G_k(x)$ и применяя стандартный алгоритм декодирования для кодов Гоппы, можно вычислить вектор ошибок по его синдромным компонентам, т. е. фактически получить вектор $\mathbf{b}_k$, а следовательно, значение информационного вектора k-го поля. Таким образом мы восстановим значение запрашиваемого поля записи.

Заключение

Рассмотрен протокол, обеспечивающий анонимность запроса к полям БД, использующий свойства вложенных кодов Гоппы.

Литература

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. **Hemenway B., Ostrovsky R.** Public Key Encryption hich is Simultaneously a Loccaly-Decodable Error-Correcting Code // Electronic Collocquium on Computational Complexity. 2007. TR07-021. http://www.eccc.uni-trier.de (дата обращения: 03.11.2010).
3. **Yekhanin S.** New Locally Decodable Codes and Private Information Retrieval Schemes // Electronic Colloquium on Computational Complexity. 2006. TR06-127. http://eccc.hpi-web.de/eccc-reports/2006/TR06-127/index.html (дата обращения: 03.11.2010).
4. **Goldreich O., Karloff H., Schulman L., Trevisan L.** Lower bounds for linear locally decodable codes and private information retrieval systems: Proc. of the 17th IEEE Conf. on Complexity Theory. IEEE Computer Society Press, 2002. P. 263–296.
5. **Bezzateev S. V., Shekhunova N. A.** On the Subcodes of one class Goppa codes: Proc. Intern. Workshop Algebraic and Combinatorial Coding Theory (ACCT-1). Sept. 1988. P. 143–146.