Защищенная БД. Обзор.

Кренделев С.Ф.
Новосибирский государственный университет
ООО "Параллелз"

Работа выполнена при финансовой поддержке Минобрнауки РФ
(договор № 02.G25.31.0054).

Безопасная база данных.

- Облако — недоверенная среда.
- Как обеспечить безопасность данных?
  - Стойкое шифрование AES или ГОСТ.
  - Проблема: невозможны типичные для БД операции на данными.
- Вывод: необходимо специальное шифрование.
- Действующий прототип:
  - CryptDB (MIT, http://css.csail.mit.edu/cryptdb/)

Проблемы создания безопасной БД.

1. Специальный вид шифрования для каждого типа данных.
2. Выбор типа шифрования в зависимости от необходимых операций:
    1. Order Preserving Encryption (OPE),
    2. Fully Homomorphic Encryption (FHE).
3. Стойкие генераторы псевдослучайных чисел (ГПСЧ).
4. Представление зашифрованных данных.

• OPE: Общий принцип.

- A, B — упорядоченные множества.
- F: A → B — отображение, сохраняющее порядок, если:
  - для любых x, y из A, если x < y, то F(x) < F(y);
  - для любых x, y из A, если F(x) < F(y), то x < y.
- Для шифрования необходимо уметь строить много разных отображений F.

OPE: Пример. Криптосистема.

- A — положительные целые числа.
- B — множество векторов $\langle a_1, a_2 \rangle$ с л-граф. порядком.
- Ключ:
  - M — положительное число;
  - S — матрица n×n из {0, 1}, det(S) = 1 mod 2;
  - Perm — случайное правило обхода всех ячеек S.
- S^k — k-я степень матрицы S по модулю 2.
- $s_k = d(S^k)$ — количество единиц в матрице S^k.

OPE: Пример. Шифрование.

- Z — исходное число. W=Z+M.
- Находим r такое, что:

  $s_1+s_2+\ldots+s_r \le W < s_1+s_2+\ldots+s_r+s_{r+1}$
- $U=W-(s_1+s_2+\ldots+s_r)$
- Находим минимальное число t такое, что на шаге t обхода матрицы S^{r+1} в порядке Perm будет встречено U единиц.
- Шифротекст: F(Z) = <r-1, t>
- Дешифрование — в обратном порядке.

FHE: Общий принцип.

- A, B — алгебраические кольца (операции +, ×).
- Отображение F: A → B — гомоморфизм, если:
  - $F(1_A) = 1_B$.
  - F(x+y) = F(x) + F(y).
  - F(x×y) = F(x) × F(y).
- FHE: построение шифрования, где шифрующая функция является гомоморфизмом.
- Подробности: «Гомоморфное шифрование. Защищенные облачные вычисления», С.Ф. Кренделев; РусКрипто 2011.

  http://goo.gl/gCGG6

ГПСЧ: Общий подход.

- Стандартные методы:
  - сдвиговый регистр с линейными обратными связями,
  - метод производящих функций.
- Обобщение:
  - решение уравнения

    $a(x)\,u(x) = b(x)$

    относительно $u(x)$ в кольце многочленов на конечным полем (или кольцом) K.

ГПСЧ: Производящие функции.

- **Определение.** Алгебраической производящей функцией называется многочлен u(x) из K[[x]], являющийся решением уравнения
- a0(x) + a1(x)u(x) + a2(x)u2(x) + … + ad(x)ud(x) = 0,
- где ai(x) — заданные многочлены над K.
- Примеры:
  - *Эллиптическая кривая*: $u^2(x)–x^3–4x–20=4$, $K=Z_{29}$.
  - *Числа Каталана*: $xu^2(x)–u(x)+1=0$.
  - *Рекуррентная последовательность* $(n+2)u_{n+2}–3(2n+3)u_{n+1}–3(n+1)u_n$: $(1–2x–3x^2)u^2(x)–1=0$.

ГПСЧ: Экспериментальные данные.

- Параметры:
  - d=2: $a_0(x) + a_1(x)\, u(x) = 0$
  - Поля Z_p, $GF(2^n)$.
- Результаты:
  - высокая энтропия генерируемых последовательностей,
  - в т. ч. для коротких контекстов.

ГПСЧ: P-адические числа.

- Решение уравнения f(x) = 0, где f(x) — многочлен с целочисленными коэффициентами, в p-адических числах.
- Последовательность имеет циклы, если ур-е:
  - имеет рациональный корень.
- Последовательность не имеет циклов, если ур-е:
  - не имеет рациональных корней,
  - неразрешимо в радикалах,
  - имеет решение только в комплексных числах.

ГПСЧ: Экспериментальные данные.

- Параметры:
  - p-адические числа.
- Результаты:
  - высокая энтропия генерируемых последовательностей,
  - в т. ч. для коротких контекстов.
- Недостатки:
  - низкая скорость генерации последовательности.

Шифрование: традиционный подход.

- Традиционный подход:
  - A — множество открытых текстов.
  - B — множество шифротекстов.
  - F_k: A → B — семейство шифрующих функций, где k — секретный ключ.
- Шифротекст — входные данные функции расшифрования, наряду с секретным ключом.

Шифрование: иной подход. Идея.

- Альтернативный подход:
  - A — поле открытых текстов.
  - k — выбранный секрет (число).
  - B — множество функций F(y), таких, что:
    - для открытого текста x, F(k) = x.
- Шифротекст — описание функции расшифрования, зависящей от секретного ключа.

Шифрование: иной подход. Пример.

- Блок шифрования: y_1, y_2, y_3, y_4 из A.
- Набор ключей: x_1, x_2, x_3, x_4.
- Шифротекст: многочлен f(x) такой, что:
  - $f(x_i) = y_i$, i=1,2,3,4.
- $f(x) = r(x)\,(x-x_1)\,(x-x_2)\,(x-x_3)\,(x-x_4) + v(x)$
  - r(x) — произвольный.
  - deg [v(x)] = 3; $v(x_i) = y_i$ — интерполируется.
  - deg [f(x)] = k.
- Представление f(x) — последовательность k+1 коэффициентов.

Шифрование: иной подход. Результаты.

- Реализован для GF(2^n).
- Недостатки:
  - шифротекст больше открытого текста,
  - медленное шифрование.
- Достоинства:
  - расшифрование ~2 порядка быстрее шифрования.

Спасибо за внимание.

Ваши вопросы?