

КРИПТОАНАЛИЗ ПО ИЗВЕСТНЫМ ОТКРЫТЫМ ТЕКСТАМ ГОМОМОРФНОЙ КРИПТОСИСТЕМЫ ДОМИНГО-ФЕРРЕ

А.В.Трепачева

Южный Федеральный Университет

4 декабря 2014 г.

Содержание

- 1 Введение в гомоморфное шифрование
- 2 Анализ по известным открытым текстам
криптосистемы Доминго-Ферре

Гомоморфное шифрование

- **Гомоморфные криптосистемы** позволяют проводить вычисления программ над зашифрованными данными.
- Результат вычислений над шифртекстами является шифровкой результата вычислений над соответствующими исходными (незашифрованными) данными.

Модель вычислений и гомоморфное шифрование

- Исходные данные – элементы некоторого кольца \mathcal{M} с операциями $+$, $*$, шифртексты – элементы кольца \mathcal{C} с операциями $+_{\mathcal{C}}$, $*_{\mathcal{C}}$
- Программа \mathbb{P} , которую нужно вычислить над вектором данных (a_1, \dots, a_n) , $a_i \in \mathcal{M}$ – набор полиномов:

$$f_j(x_1, \dots, x_n) = \sum_{\{i_1, \dots, i_t\} \in \{1, \dots, n\}} f_{i_1, \dots, i_t}^j * x_{i_1} * \dots * x_{i_t}, f_{i_1, \dots, i_t}^j \in \mathcal{M}, j = 1..m$$

- Вычисление программы:
 $b_j \in \mathcal{M} \leftarrow f_j(a_1, \dots, a_n), j = 1..m.$
- Тогда криптосистема гомоморфна, если для $\forall x, y \in \mathcal{M}$:
 $D(E(x) +_{\mathcal{C}} E(y)) = x + y, D(E(x) *_{\mathcal{C}} E(y)) = x * y$, где E, D – функции зашифрования и расшифрования.

Где нужно гомоморфное шифрование?

Потребность в гомоморфном шифровании возникает в случае, когда вычисления над приватными данными проводит Вычислитель, которому Владелец данных не доверяет.

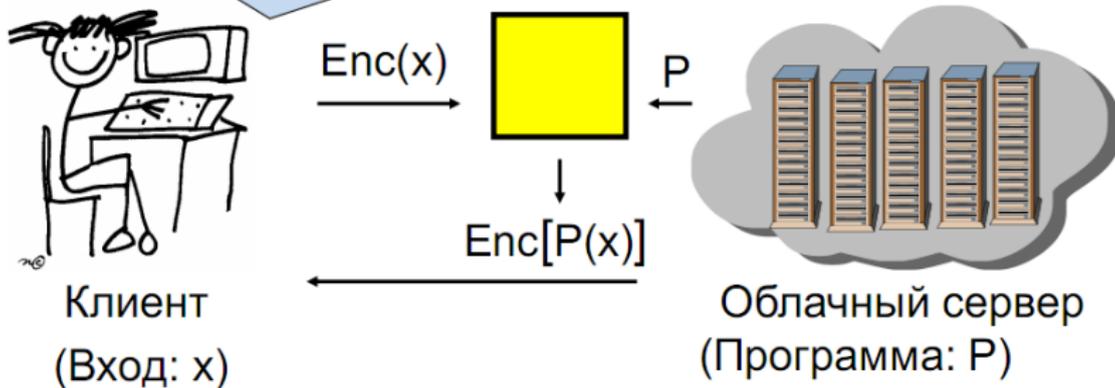
Такая ситуация возникает, например:

- *В Облачных сервисах.*
- *При передаче данных по Беспроводным Децентрализованным Сетям Связи (Mobile Ad hoc Network, MANET).*

Гомоморфное шифрование и Облачные вычисления

Клиент хочет делегировать вычисление облачному серверу, однако при этом он не желает, чтобы его входные данные попали на сервер в открытом виде.

Поэтому нужна криптосистема, которая позволила бы провести вычисления над шифртекстами с последующей возможностью извлечь за приемлемое время из результирующей шифровки результат нужных вычислений над открытыми данными.



Цель работы

- В данной работе уточняется степень защищенности гомоморфной криптосистемы Доминго-Ферре против атаки по известным открытым текстам.
- Почему анализ криптосистемы Доминго-Ферре представляет интерес?

Эту криптосистему можно эффективно использовать в некоторых приложениях (например её используют для защиты данных в Mobile Ad hoc Networks).

Однако, шифруя с её помощью данные, нужно хорошо осознавать насколько она уязвима к некоторым атакам.

Симметричная гомоморфная криптосистема Доминго-Ферре

- Открытый текст** a – элемент кольца вычетов \mathbb{Z}_n ,
 $n = p \cdot q$, где p, q – большие простые числа такие, что
 $p < q$, $\log_2(p) \approx \log_2(q)$.
 n – публичное, а p, q являются секретом.
- Ключ** – $k = (k_p \in \mathbb{Z}_p \setminus \{0\}, k_q \in \mathbb{Z}_q \setminus \{0\})$.
- Шифртекст** $c = \{c_p(x), c_q(x)\} \in \mathbb{Z}_n[x] \times \mathbb{Z}_n[x]$, где
 $c_p(0) = 0$, $c_q(0) = 0$, $\deg(c_p(x)) = d$, $\deg(c_q(x)) = d$,
 $d \in \mathbb{Z}_+$ – параметр.

$$c_p(k_p) \bmod p = a_p \in \mathbb{Z}_p (\equiv a \pmod{p})$$

$$c_q(k_q) \bmod q = a_q \in \mathbb{Z}_q (\equiv a \pmod{q})$$
- Тогда по *Китайской теореме об остатках*:

$$a = q \cdot (q^{-1} \bmod p) \cdot a_p + p \cdot (p^{-1} \bmod q) \cdot a_q.$$

Уточнение процедуры шифрования

Шифрование ($a \in \mathbb{Z}_n, d \in \mathbb{Z}_+, \mathbf{p}, \mathbf{q}, k_p \in \mathbb{Z}_p \setminus \{0\}, k_q \in \mathbb{Z}_q \setminus \{0\}$)

- $a \in \mathbb{Z}_n \longrightarrow a'(x) = \sum_{i=1}^d a'_i \cdot x^i \in \mathbb{Z}_n[x]$, где:

$$2) a'_d \stackrel{\$}{\leftarrow} \mathbb{Z}_n \setminus \{0\}$$

2) $a'_i \stackrel{\$}{\leftarrow} \mathbb{Z}_n, i = \overline{2, d-1}$ (a'_i берутся по равномерному распределению над \mathbb{Z}_n)

$$3) a'_1 := (a - \sum_{i=2}^d a'_i) \bmod n$$

Тогда имеем:

$$a \equiv \sum_{i=1}^d a'_i \pmod{n}$$

- Шифртекст $c = \{c_p(x), c_q(x)\} \in \mathbb{Z}_n[x] \times \mathbb{Z}_n[x]$, где:

$$c_p(x) := a'(k_p^{-1} \cdot x) \bmod \mathbf{p}, \text{ и } \Rightarrow c_p(k_p) \equiv a \pmod{\mathbf{p}}$$

$$c_q(x) := a'(k_q^{-1} \cdot x) \bmod \mathbf{q}, \text{ и } \Rightarrow c_q(k_q) \equiv a \pmod{\mathbf{q}}$$

То есть на самом деле все коэффициенты $c_p(x) \in \mathbb{Z}_n[x]$ принадлежат $\{0, \dots, p-1\}$, коэффициенты $c_q(x) \in \mathbb{Z}_n[x]$ принадлежат $\{0, \dots, q-1\}$.

Где гомоморфизм?

Пусть даны $a_i \in \mathbb{Z}_n, i = 1, 2$ и их шифровки $c_i = \{c_{p,i}(x), c_{q,i}(x)\}, i = 1, 2$ на одном ключе $k = (k_p, k_q)$.

По свойствам сравнений имеем:

- **Сумма:** $c_+ = \{(c_{p,1}(x) + c_{p,2}(x)) \bmod n, (c_{q,1}(x) + c_{q,2}(x)) \bmod n\}$ шифрует $(a_1 + a_2) \bmod n$ на ключе k .
- **Произведение:**
 $c_* = \{(c_{p,1}(x) \cdot c_{p,2}(x)) \bmod n, (c_{q,1}(x) \cdot c_{q,2}(x)) \bmod n\}$ шифрует $(a_1 \cdot a_2) \bmod n$ на ключе k .

Основное свойство секретного ключа

$$a \in \mathbb{Z}_n \longrightarrow c = \{c_p(x), c_q(x)\} \in \mathbb{Z}_n[x] \times \mathbb{Z}_n[x]$$

$$c_p(k_p) \bmod \mathbf{p} \equiv a \pmod{\mathbf{p}}$$

$$c_q(k_q) \bmod \mathbf{q} \equiv a \pmod{\mathbf{q}}$$

То есть имеем:

- k_p – корень $f_p(x) = c_p(x) - a \in \mathbb{Z}_n[x]$ по модулю p .
- k_q – корень $f_q(x) = c_q(x) - a \in \mathbb{Z}_n[x]$ по модулю q .
- Но $f_p(k_p) \not\equiv a \pmod{n}$, $f_q(k_q) \not\equiv a \pmod{n}$.

План криптоанализа по известным открытым текстам

Пусть у криптоаналитика есть t пар:

$(a_i \in \mathbb{Z}_n, c_i = \{c_{p,i}(x), c_{q,i}(x)\} \in \mathbb{Z}_n[x] \times \mathbb{Z}_n[x]), i = \overline{1, t}$, где c_i шифрует a_i .

Причем все c_i изготовлены на одном ключе $k = (k_p, k_q)$ и $\forall i : \deg(c_{p,i}(x)) = d, \deg(c_{q,i}(x)) = d$.

Для того чтоб взломать криптосистему криптоаналитику нужно:

- Раскрыть факторизацию n , т.е. найти p, q .
- Вычислить k_p как общий корень $f_{p,i}(x) = c_{p,i}(x) - a_i$ по модулю p .
- Вычислить k_q как общий корень $f_{q,i}(x) = c_{q,i}(x) - a_i$ по модулю q .

Известные результаты по криптоанализу

Ранее в литературе был представлен результат состоящий в том, что для того, чтобы раскрыть p, q и $k = (k_p, k_q)$ при знании $n = p \cdot q$ криптоаналитику необходимо перехватить $t \geq d + 1$ пар (открытый текст, шифртекст), изготовленных на ключе k .

Существующий метод криптоанализа (вычисление \mathbf{p})

Пусть $\mathbf{t} = \mathbf{d} + \mathbf{1}$, даны $\mathbf{f}_{\mathbf{p},i}(x) = c_{\mathbf{p},i}(x) - a_i = \sum_{j=1}^d c_{\mathbf{p},i,j} \cdot x^j - a_i$
 $\in \mathbb{Z}_n[x]$, $i = \mathbf{1}, \dots, \mathbf{d} + \mathbf{1}$.

$$\mathbf{A} = \begin{pmatrix} -a_1 & c_{\mathbf{p},1,1} & \cdots & c_{\mathbf{p},1,d} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{\mathbf{d}+1} & c_{\mathbf{p},\mathbf{d}+1,1} & \cdots & c_{\mathbf{p},\mathbf{d}+1,d} \end{pmatrix} \quad (1)$$

- СЛАУ $(\mathbf{A}|\mathbf{0})$ имеет решение $(1, k_{\mathbf{p}}, k_{\mathbf{p}}^2, \dots, k_{\mathbf{p}}^d)$ по модулю $\mathbf{p} \Rightarrow \det(\mathbf{A}) \bmod \mathbf{p} = \mathbf{0}$ и

$\Rightarrow \det(\mathbf{A}) \bmod \mathbf{n} = \mathbf{p} \cdot \mathbf{k}$, $\mathbf{k} \in \{\mathbf{0}, \dots, \mathbf{q} - \mathbf{1}\}$.

Если $\mathbf{k} \neq \mathbf{0}$, то НОД $(\det(\mathbf{A}) \bmod \mathbf{n}, \mathbf{n}) = \mathbf{p}$.

- $\mathbf{k} = \mathbf{0} \iff \det(\mathbf{A}) \bmod \mathbf{q} = \mathbf{0}$.
- Если \mathbf{A} – матрица с \approx равномерно случайными элементами по модулю \mathbf{q} , то $\Pr(\det(\mathbf{A}) \bmod \mathbf{q} \neq \mathbf{0}) \approx \mathbf{1}$ (для большого \mathbf{q}).

Однако \mathbf{A} будет таковой, если вероятностное распределение \mathcal{D} на открытых текстах будет равномерным (см. 1-й столбец \mathbf{A}).

Существующий метод криптоанализа (вычисление k_p)

Даны $f_{p,i}(x) = c_{p,i}(x) - a_i = \sum_{j=1}^d c_{p,i,j} \cdot x^j - a_i$
 $\in \mathbb{Z}_n[x]$, $i = 1, \dots, d + 1$.

- $f_{p,i}(k_p) \bmod p = 0$ и \Rightarrow

$$f_{p,i}(x) \equiv (x - k_p) \cdot g_i(x) \pmod{p},$$

$g_i(x) \in \mathbb{Z}_p[x]$ – равномерно случайные.

- $\Rightarrow \text{НОД}(f_{p,1}(x), \dots, f_{p,d+1}(x)) \equiv x - k_p \pmod{p}$
 с вероятностью:

$$1 - \frac{1}{p^{d+1}} + \frac{p-1}{p^{(d+1)^2}}.$$

Результант полиномов

Для $\forall f(x), g(x) \in \mathbb{Z}_n[x]$, $n = p \cdot q$, можно составить матрицу Сильвестра $S \in \mathbb{Z}_n^{(d_1+d_2) \times (d_1+d_2)}$, где $\deg(f) = d_1$, $\deg(g) = d_2$:

$$S = \begin{pmatrix} f_0 & \cdots & f_{d_1} & 0 & 0 & \cdots & 0 \\ 0 & f_0 & \cdots & f_{d_1} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & 0 & f_0 & \cdots & f_{d_1} \\ g_0 & \cdots & g_{d_2} & 0 & 0 & \cdots & 0 \\ 0 & g_0 & \cdots & g_{d_2} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & 0 & g_0 & \cdots & g_{d_2} \end{pmatrix} \quad (2)$$

Результант $f(x), g(x) - \text{Res}(f, g) := \det(S) \bmod n \in \mathbb{Z}_n$.

Если $f(x), g(x)$ имеют общий корень по модулю n , то $\text{Res}(f, g) = 0$.

Если $f(x), g(x)$ имеют общий корень по модулю p или q , то $\text{Res}(f, g) \equiv 0$ по модулю p или q .

Вычисление p по двум парам

Дано:

$(a_i \in \mathbb{Z}_n, c_i = \{c_{p,i}(x), c_{q,i}(x)\} \in \mathbb{Z}_n[x] \times \mathbb{Z}_n[x]), i = 1, 2.$

$f_{p,1}(x) = c_{p,1}(x) - a_1, f_{p,2}(x) = c_{p,2}(x) - a_1 \in \mathbb{Z}_n[x]$

- Так как $f_{p,1}(k_p) \equiv 0 \pmod{p}$ и $f_{p,2}(k_p) \equiv 0 \pmod{p}$, то

$$\mathcal{R} = \text{Res}(f_{p,1}(x), f_{p,2}(x)) \equiv 0 \pmod{p}.$$

- Тогда $\mathcal{R} = p \cdot k$, где $k \in \{0, 1, 2, \dots, q - 1\}$.
- Если $\mathcal{R} \neq 0$, то $p = \text{GCD}(n, \mathcal{R})$.
- $\mathcal{R} = 0 \iff \mathcal{R} \bmod q = 0$ (по Китайской теореме об остатках).

Какова вероятность, что $\mathcal{R} \bmod \mathbf{q} \neq \mathbf{0}$?

$$\mathbf{f}_{p,i}(x) = c_{p,i}(x) - a_i = \sum_{j=1}^d c_{p,i,j} \cdot x^j - a_i \in \mathbb{Z}_n[x], i = 1, 2$$

- Если предположить, что $\mathbf{f}_{p,i}^*(x) = \mathbf{f}_{p,i}(x) \bmod \mathbf{q}, i = 1, 2$ – равномерно случайные полиномы $\in \mathbb{Z}_q[x]$, то

$$Pr\{\mathcal{R} \bmod \mathbf{q} \neq \mathbf{0}\} = 1 - \frac{1}{\mathbf{q}}.$$

- Однако на самом деле $\mathbf{f}_{p,i}^*(x) = \sum_{j=0}^d \mathbf{f}_{i,j} \cdot x^j, i = 1, 2$ не будут равномерно случайными, если вероятностное распределение \mathcal{D} на пространстве открытых текстов \mathbb{Z}_n неравномерно.
- И даже если \mathcal{D} равномерно, то $\mathbf{f}_{p,i}^*(x), i = 1, 2$ строго говоря не будут равномерно случайными в $\mathbb{Z}_q[x]$, поскольку в соответствии с процедурой шифрования: $0 < c_{p,i,j} < p$.

Так что: $\mathbf{f}_{i,0} \stackrel{\$}{\leftarrow} \{0, \dots, \mathbf{q} - 1\}, \mathbf{f}_{i,j} \stackrel{\$}{\leftarrow} \{0, \dots, \mathbf{p} - 1\}$ для $1 \leq j \leq d$.

Практическая оценка вероятности найти p, q

Вероятностное распределение \mathcal{D} на пространстве открытых текстов \mathbb{Z}_n полагалось равномерным.

Pr_1 – практическая оценка вероятности, того, что $\mathcal{R} \bmod q \neq 0$ (при этом для практической оценки вероятности поводилось $\approx 10^5$ итераций.)

$$\text{Pr}_2 = 1 - 1/q$$

Таблица: Для $d = 10$

n	p	q	Pr_1	Pr_2
6	2	3	0.67	0.67
35	5	7	0.86	0.86
91	7	13	0.92271	0.923
253	11	23	0.956	0.957
1517	37	41	0.97	0.97
3599	59	61	0.98	0.98
9991	97	103	0.99	0.991

Практическая оценка вероятности найти p, q

Вероятностное распределение \mathcal{D} на пространстве открытых текстов \mathbb{Z}_n полагалось равномерным.

Pr_1 – практическая оценка вероятности, того, что $\mathcal{R} \bmod q \neq 0$ (при этом для практической оценки вероятности поводилось $\approx 10^5$ итераций.)

$$\text{Pr}_2 = 1 - 1/q$$

Таблица: Для $d = 50$

n	p	q	Pr_1	Pr_2
15	3	5	0.8	0.8
221	13	17	0.92	0.94
1147	31	37	0.964	0.972
2173	41	53	0.999	0.999
13943	103	131	0.999	0.999

Вычисление k_p, k_q

Даны: $f_{p,i}(x) = c_{p,i}(x) - a_i$, $f_{q,i}(x) = c_{q,i}(x) - a_i$, $i = 1, 2$

- $f_{p,i}(k_p) \bmod p = 0$, $i = 1, 2$ и \Rightarrow

$$f_{p,i}(x) \equiv (x - k_p) \cdot g_{p,i}(x) \pmod{p}, i = 1, 2,$$

$g_{p,i}(x) \in \mathbb{Z}_p[x]$ – равномерно случайные (в соответствии с процедурой шифрования).

- \Rightarrow НОД $(g_{p,1}(x), g_{p,2}(x)) \equiv 1 \pmod{p}$ с вероятностью $1 - 1/p$.
- Итого НОД $(f_{p,1}(x), f_{p,2}(x)) \equiv x - k_p \pmod{p}$ с вероятностью $1 - 1/p$.
- Аналогично НОД $(f_{q,1}(x), f_{q,2}(x)) \equiv x - k_q \pmod{q}$ с вероятностью $1 - 1/q$.

Оценка вероятности раскрыть k_p

\Pr_1 – практическая оценка вероятности, того, что $\text{НОД}(f_1(x), f_2(x)) = 1$, где $f_1(x), f_2(x)$ – равномерно случайные полиномы из $\mathbb{Z}_p[x]$ (для практической оценки вероятности поводилось $\approx 10^5$ итераций.)

$$\Pr_2 = 1 - 1/p$$

Таблица:

p	\Pr_1	\Pr_2
2	0.48	0.5
23	0.7	0.7
103	0.99	0.9902
3389	1	1

Асимптотические оценки сложности криптоанализа

- Вычисление \mathbf{p}, \mathbf{q} потребует не более, чем $O(d^3 \cdot \log_2^2(n))$ операций по модулю \mathbf{n} .
- Вычисление k_p потребует соответственно $O(d^2 \cdot \log_2^2(p))$ операций по модулю \mathbf{p} .
- Вычисление k_q потребует соответственно $O(d^2 \cdot \log_2^2(q))$ операций по модулю \mathbf{q} .

Время работы атаки

T – время затраченное на раскрытие p, q и $k = (k_p, k_q)$
 $\log_2(n) = 1024, \log_2(p) = 512, \log_2(q) = 512$

Таблица:

d	T
4	14 мсек
8	38 мсек
16	121 мсек
32	460 мсек
64	1.9 сек
128	9.5 сек
256	52 сек
512	5 мин
1024	22 мин

Для реализации использовались Qt 1.3.1 и библиотека NTL.
Замеры времени проводились на ПК со следующими характеристиками: Quad Core Celerone 1.7 GHz with 4 GB

Время работы атаки

T – время затраченное на раскрытие p, q и $k = (k_p, k_q)$.

$\log_2(n) = 2048, \log_2(p) = 1024, \log_2(q) = 1024$

Таблица:

d	T
4	39 мсек
8	112 мсек
16	387 мсек
32	1.5 сек
64	6 сек
128	27 сек
256	2 мин
512	12 мин
1024	50 мин

Заключение

- Итак, мы получили, что с вероятностью ≈ 1 криптосистему Доминго Ферре можно взломать при наличии только двух пар (открытый текст, шифртекст).
- Однако важно отметить, что как наша, так и предложенная ранее атаки будут гарантированно работать если вероятностное распределение \mathcal{D} на открытых текстах равномерное.
- Если \mathcal{D} такое, что $Pr(0)$ не пренебрежительно мала, то вероятность успеха криптоанализа сильно падает.
- В случае если \mathcal{D} неравномерно, но при этом $Pr(0) \ll 1$, нужно дополнительное исследование.
- В дальнейшем планируется провести более подробный криптоанализ криптосистемы Доминго Ферре только по шифртекстам

Спасибо за внимание!