Гомоморфное шифрование [1]

Н.П. Варновский, А.В. Шокуров

Аннотация. Гомоморфное шифрование как криптографический примитив представляет интерес как с прикладной, так и с чисто математической точек зрения. Несмотря на многолетние исследования в этой области, основные проблемы остаются нерешенными. Настоящая статья является учебно-методической. Не претендуя на полноту обзора результатов, относящихся к гомоморфному шифрованию, она знакомит читателя с данным криптографическим примитивом и некоторыми смежными вопросами.

1. Введение

В литературе по теоретической криптографии под гомоморфным шифрованием понимается криптографический примитив, представляющий собой функцию шифрования, удовлетворяющую дополнительному требованию гомоморфности относительно каких-либо алгебраических операций над открытыми текстами.

Пусть $E(k,m)$ — функция шифрования, где m — открытый текст, k — ключ шифрования. Заметим, что для данных фиксированных k и m криптограмма $E(k,m)$ может быть, вообще говоря, случайной величиной. В таких случаях говорят о вероятностном шифровании. Функция E гомоморфна относительно операции op над открытыми текстами, если существует эффективный алгоритм M, который получив на вход любую пару криптограмм вида $E(k,m_1)$, $E(k,m_2)$, выдает криптограмму c такую, что при дешифровании c будет получен открытый текст $m_1 op m_2$.

Как правило, рассматривается следующий важнейший частный случай гомоморфного шифрования. Для данной функции шифрования E и операции op_1 над открытыми текстами существует операция op_2 над криптограммами такая, что из криптограммы $E(k,m_1) op_2 E(k,m_2)$ при дешифровании извлекается открытый текст $m_1 op_1 m_2$.

[1] Работа выполнена при поддержке РФФИ, проект 06-01-00106.

Алгоритм M, вообще говоря, вероятностный. Существует специальная модификация рассматриваемого примитива, называемая гомоморфным шифрованием с рерандомизацией. В таком случае при фиксированных $E(k,m_1)$ и $E(k,m_2)$ криптограмма c — случайная величина, причем требуется, чтобы по заданным c, $E(k,m_1)$, $E(k,m_2)$, но при неизвестном ключе дешифрования, было невозможно эффективно проверить, что криптограмма c получена из $E(k,m_1)$ и $E(k,m_2)$, т. е. содержит открытый текст $m_1\ op\ m_2$.

Гомоморфное шифрование как криптографический примитив может найти широкое применение в криптографии и, в более широком смысле, в разработке математических методов защиты информации. Здесь прежде всего следует выделить такую, интересную с прикладной точки зрения, задачу как вычисления над зашифрованными данными. Конфиденциальные данные хранятся в зашифрованном виде. Для выполнения вычислений над ними данные можно расшифровать, произвести необходимые операции, и затем результаты вновь зашифровать. Но для этого требуются защищенная аппаратура и, уж по крайней мере, организационные меры по хранению секретных ключей. Вычисления над зашифрованными данными, если они возможны, помогают избежать всех этих проблем.

Задачу вычисления над зашифрованными данными можно рассматривать в различных постановках. Например, данными может быть массив натуральных чисел $a_1,\ldots,a_n$ из диапазона от 1 до N. Предположим, что система вычисления над зашифрованными данными позволяет противнику выполнить следующий запрос. Выбираются произвольное число a из того же диапазона от 1 до N и индекс $i \in \{1,\ldots,n\}$. На запрос (a,i) противник получает ответ 0, если $a \geq a_i$, и ответ 1, если $a < a_i$. Очевидно, что какой бы высокой стойкостью не обладала криптосистема, противник всегда может, используя метод деления пополам, определить значение числа a_i. Таким образом, в данной постановке вычисления над зашифрованными данными невозможны.

Наиболее популярный из рассматриваемых в литературе вариантов системы вычислений над зашифрованными данными не включает в себя операции сравнения. Реализуется лишь некоторый „полный“ набор операций. В случае, когда операндами являются биты, это может быть, например, базис И, ИЛИ, НЕ, а для произвольных числовых операндов — операции сложения и умножения. Ясно, что такую систему вычислений над зашифрованными данными можно было бы легко реализовать, если бы существовала функция шифрования, гомоморфная сразу относительно двух операций: И и ИЛИ в случае булевых

операндов, сложение и умножение в случае числовых. Однако, вопрос о существовании таких функций гомоморфного шифрования, равно как и систем вычислений над зашифрованными данными, остается открытым.

Легко видеть, что как гомоморфное шифрование, так и вычисления над зашифрованными данными можно реализовать, исходя из предположения о существовании стойкой обфускации. В самом деле, пусть P' — программа, выполняющая вычисления над данными x. Преобразуем ее в новую программу P, которая содержит в себе ключ дешифрования и получает на вход зашифрованные данные $E(k, x)$. Программа P выполняет дешифрование и вызывает P' как подпрограмму, передавая ей на вход данные x. После завершения работы подпрограммы P' программа P шифрует результат и останавливается. Пусть $\mathcal{O}(\mathcal{P})$ — стойкая обфускация программы P. Ясно, что $\mathcal{O}(\mathcal{P})$ обеспечит вычисления, эквивалентные вычислениям программы P', но при этом данные останутся недоступными противнику, т. е., будет достигнут тот же эффект, что при вычислении над зашифрованными данными. Подчеркнем, что для того, чтобы данная методика работала, необходима обфускация $\mathcal{O}$, стойкая в смысле Барака и др. [5].

Можно предположить, что верна и обратная импликация: если существует подходящая функция гомоморфного шифрования, то существует и стойкая обфускация. Интуитивные соображения в поддержку этого предположения таковы. Пусть U — универсальная программа, которая, получив на вход текст произвольной программы P и входные данные x, выполняет вычисления программы P на этом входе. Далее, пусть E — функция шифрования, гомоморфная относительно всех операций, используемых программой U. Тогда пару $(E(k, P), U)$ можно рассматривать как обфускацию программы P. Здесь для простоты мы обозначили универсальную программу, выполняющую вычисления над зашифрованными данными, той же буквой U.

Вопрос о соотношении между предположениями о существовании функций гомоморфного шифрования различных типов и предположениями о существовании стойких обфускаторов остается по существу открытым.

2. Гомоморфные системы шифрования

2.1. Криптосистема RSA

Криптосистема RSA является самой популярной криптографической схемой, описание которой можно найти практически в любом тексте, посвященном криптографии с открытым ключом. Поэтому здесь мы эту криптосистему не описываем, ограничимся обозначениями. Пусть N — составной модуль (модуль RSA) и e — открытая экспонента. Таким образом, пара (N, e) является открытым ключом криптосистемы. Далее, пусть $m \in Z_N$ — открытый текст. Функция шифрования $E((N, e), m) = m^e \bmod N$ криптосистемы RSA гомоморфна относительно умножения открытых текстов. В самом деле, для любых двух открытых текстов m_1, m_2 и любого открытого ключа k криптограмма произведения равна произведению криптограмм сомножителей: $E(k, m_1 \cdot m_2) = E(k, m_1) \cdot E(k, m_2)$.

2.2. Криптосистема Эль Гамаля

Пусть G — циклическая группа порядка p и g — ее порождающий. В качестве секретного ключа криптосистемы выбирается случайный элемент x группы Z_{p-1}. Соответствующий открытый ключ вычисляется по формуле $y = g^x$. Криптограмма открытого текста $m \in G$ вычисляется с помощью функции шифрования $E(y, m) = (y^r m, g^r)$, где r — случайный элемент группы Z_{p-1}, т. е. число r выбирается всякий раз заново, независимо и равновероятно.

Дешифрование криптограммы (c_1, c_2) выполняется следующим образом. Сначала вычисляется $c_2^x = g^{rx}$, откуда $m = c_1/c_2^x$.

Криптосистема Эль Гамаля является криптосистемой вероятностного шифрования. Ее функция шифрования гомоморфна относительно операции умножения открытых текстов: криптограмма произведения может быть вычислена как произведение (попарное) криптограмм сомножителей. Если $E(y, m_1) = (y^{r_1} m_1, g^{r_1})$ и $E(y, m_2) = (y^{r_2} m_2, g^{r_2})$, то $E(y, m_1 m_2)$ можно получить в виде $(y^{r_1} y^{r_2} m_1 m_2, g^{r_1} g^{r_2})$.

Функция шифрования криптосистемы Эль Гамаля обладает свойством рерандомизации. Криптограмму произведения, полученную в указанном выше виде, можно рандомизировать, выбрав случайное число r из Z_{p-1} и домножив первый элемент криптограммы на y^r, а второй — на g^r.

Таким образом будет получена криптограмма $(y^{r_1} y^{r_2} y^r m_1 m_2, g^{r_1} g^{r_2} g^r)$, связь которой с исходными криптограммами „затемнена“ случайными множителями.

2.3. Гомоморфное шифрование в билинейных группах

В работе [1] описана схема шифрования в группе с билинейным спариванием, гомоморфная относительно произвольного числа групповых операций и одного билинейного спаривания.

Опишем основные идеи данной конструкции. Вначале определим понятие билинейной группы.

Определение. Пусть $\mathbb{G}$ и $\mathbb{G}_1$ — две (мультипликативные) циклические группы конечного порядка n, и g — образующий группы $\mathbb{G}$. Отображение

$$e : \mathbb{G} \times \mathbb{G} \mapsto \mathbb{G}_1$$

называется билинейным, если выполняются равенства $e(u^a, v^b) = e(u, v)^{ab}$ для всех $u, v \in \mathbb{G}$ и $a, b \in \mathbb{Z}$ и элемент $e(g, g)$ — образующий группы $\mathbb{G}_1$. Группа $\mathbb{G}$ называется билинейной, если существует группа $\mathbb{G}_1$ и билинейное отображение $e : \mathbb{G} \times \mathbb{G} \mapsto \mathbb{G}_1$.

Опишем конструкцию билинейных групп порядка n. Пусть $n > 3$ — целое число, не содержащее кратных простых делителей и не делящееся на 3. Построим билинейную группу $\mathbb{G}$ порядка n следующим образом:

1. Найдем наименьшее натуральное $l \in \mathbb{Z}$ такое, что $p = ln - 1$ — простое и $p \equiv 2 \mod 3$.
2. Рассмотрим группу точек эллиптической кривой γ, заданной уравнением $y^2 = x^3 + 1$ над полем $\mathbb{F}_p$. Поскольку $p \equiv 2 \mod 3$, эллиптическая кривая содержит $p + 1 = ln$ точек. Поэтому она содержит подгруппу порядка n. Обозначим эту подгруппу через $\mathbb{G}$.
3. Пусть $\mathbb{G}_1$ — подгруппа группы $\mathbb{F}^*_{p^2}$ порядка n. Тогда модифицированное спаривание Вейля на эллиптической кривой γ задает билинейное отображение $e : \mathbb{G} \times \mathbb{G} \mapsto \mathbb{G}_1$.

Пусть задано случайное число $\tau \in \mathbb{Z}$. Находим два простых числа p_1 и p_2, имеющих τ разрядов в двоичном представлении. Положим $n = p_1 p_2$ и построим билинейную группу $\mathbb{G}$ с помощью описанной выше конструкции. Выберем две случайных образующих $g, u \in \mathbb{G}$. Положим $h = u^{p_2}$. Тогда h — случайная образующая подгруппы группы $\mathbb{G}$ порядка p_1.

Открытым ключом тогда объявляем набор $(n, \mathbb{G}, \mathbb{G}_1, e, g, h)$. Секретным ключом является p_1.

Шифрование. Шифруются целочисленные данные в диапазоне $[0, T]$, где $T < p_2$. Криптограмма для m определяется формулой $C = g^m h^r \in \mathbb{G}$, где $r \in \{0, 1, \ldots, n-1\}$ — случайное число.

Дешифрование выполняется по формуле $\log_{g^{p_1}} C^{p_1} = \log_{g^{p_1}} (g^m h^r)^{p_1} = \log_{g^{p_1}} (g^{p_1})^m = m$. Дискретный логарифм по основанию g^{p_1} находится λ-методом Полларда за время $\tilde{O}(\sqrt{T})$, поскольку $0 \le m \le T$.

Введенное гомоморфное шифрование позволяет вычислять квадратичные формы в $\mathbb{Z}/n\mathbb{Z}$. В частности можно вычислять булевы формы вида

$$\varphi(x_1, \ldots, x_n) = \bigvee_{i=1}^{k} (l_{i,1} \wedge l_{i,2}),$$

где $l_{i,k}$ — булева формула, содержащая только операции $\vee$ и $\neg$.

2.4. Смешанное шифрование

Смешанное шифрование обобщает линейное шифрование и шифрование системой остатков. А именно, пусть $p_1, \ldots, p_m$ — взаимно простые числа. Целое $0 \le x < p_1 \cdot \ldots \cdot p_m$ как элемент кольца $\mathbb{Z}_{p_1 \cdot \ldots \cdot p_m}$ представляется при смешанном шифровании системой целых чисел с помощью формул

$$\begin{aligned} x'_1 &\equiv a_1 \cdot x_1 + b_1 \mod p_1 \\ &\ldots \\ x'_m &\equiv a_m \cdot x_m + b_m \mod p_m \end{aligned} \qquad (1)$$

где $\mathrm{GCD}(a_k, p_k) = 1$ для всех $k = 1, \ldots, m$. Ключами шифрования в этом случае являются наборы $(a_1, b_1, \ldots, a_m, b_m)$.

В работе [2] определено понятие однородной схемы вычисления и показано, что эта система шифрования является гомоморфной для таких схем.

2.5. Криптосистема Пэйе (Paillier)

Пусть p и q — два больших простых числа и пусть $n = pq$ и $\lambda = \mathrm{НОК}(p-1, q-1)$. Выберем случайное число g из $Z^*_{n^2}$ и вычислим $\mu = (L(g^\lambda \bmod n^2))^{-1} \bmod n$, где $L(u) = (u-1)/u$. Поскольку при выбранном λ выполняется сравнение $u \equiv 1 \bmod n$, определение корректно.

Открытым ключом криптосистемы служит пара $k = (n, g)$, а секретным ключом — пара (λ, μ).

Для шифрования открытого текста $m \in Z_n$ выбирается случайное число r из $Z^*_{n^2}$ и вычисляется шифртекст по формуле $c = g^m r^n \bmod n^2$.

Дешифрование криптограммы c выполняется по формуле

$$m = L(c^\lambda \bmod n^2)\mu \bmod n.$$

Тот факт, что последняя формула и в самом деле дает открытый текст, требует доказательства [4].

Криптосистема Пэйе также является системой вероятностного шифрования. Из всех известных криптосистем с открытым ключом данная обладает, по-видимому, наиболее интересными гомоморфными свойствами:

- произведение двух криптограмм является криптограммой суммы соответствующих открытых текстов, т. е. при дешифровании криптограммы $E(k, m_1) \cdot E(k, m_2) \bmod n^2$ будет получен открытый текст $m_1 + m_2 \bmod n$;
- ту же самую сумму можно получить, умножив криптограмму $E(k, m_1)$ на g^{m_2}, т. е. при дешифровании криптограммы $E(k, m_1) \cdot g^{m_2} \bmod n^2$ будет получен открытый текст $m_1 + m_2 \bmod n$;
- открытый текст, содержащийся в криптограмме, можно умножить на константу d, возведя эту криптограмму в степень d, т. е. при дешифровании криптограммы $E(k, m)^d \bmod n^2$ будет получен открытый текст $dm \bmod n$. В частности, в качестве константы d можно задать другой открытый текст m' и тем самым получить криптограмму произведения $mm' \bmod n$.

Подчеркнем, что последнее свойство не является свойством гомоморфности функции шифрования относительно операции умножения открытых текстов.

3. Пример: решение системы линейных уравнений, с зашифрованной правой частью

Рассмотрим систему линейных уравнений

$$\begin{cases} a_{11}x_1 + \ldots + a_{1n}x_n & = & b_1 \\ \ldots\ldots\ldots\ldots\ldots\ldots & \ldots & \ldots \\ a_{m1}x_1 + \ldots + a_{mn}x_n & = & b_m \end{cases}.$$

Пусть матрица системы задана в открытом виде, а правая часть в виде зашифрованного текста с использованием криптограммы Пэйе (см.п.2.5).

Будем решать эту систему уравнений методом Гаусса. Переставим строки так, чтобы первые k строк стали линейно независимыми, а остальные выражались бы через них. Зашифрованные данные в правой части также подвергнутся этой перестановке.

Рассмотрим первый столбец матрицы системы. Найдем наибольший общий делитель d_1 его элементов. Без ограничения общности будем считать, что все элементы матрицы системы взаимно просты с p и q. Поскольку p и q большие числа, вероятность противоположного события пренебрежимо мала.

Тогда используя элементарные преобразования умножения строк матрицы на число и сложения строк, приведем матрицу системы к виду

$$\begin{pmatrix} 1 & a'_{12} & \cdots & a'_{1n} \\ 0 & a'_{22} & \cdots & a'_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & a'_{m2} & \cdots & a'_{mn} \end{pmatrix}.$$

При таком преобразовании матрицы системы над элементами в правой части будут выполняться также операции сложения и умножения на незашифрованные константы. Поскольку шифрование Пэйе гомоморфно относительно этих операций, то все вычисления могут быть выполнены в зашифрованном виде и после такого преобразования матрицы правая часть также будет представлена в зашифрованном виде.

Повторим последнюю процедуру еще $k - 1$ раз. Получим матрицу вида

$$\begin{pmatrix} 1 & a'_{12} & \cdots & \cdots & a'_{1k} & \cdots & a'_{1n} \\ 0 & 1 & \cdots & \cdots & a'_{2k} & \cdots & a'_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 1 & \cdots & a'_{kn} \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Теперь матрица легко приводится к „диагональному“ виду

$$\begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 & \cdots & c_{1n} \\ 0 & 1 & \cdots & \cdots & 0 & \cdots & c_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 1 & \cdots & c_{kn} \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Заметим, что правая часть системы преобразуется с помощью формул

$$b'_i = L_i(b_1, \ldots, b_m),$$

где L_i — линейные формы, коэффициенты которых заданы в незашифрованном виде. Поскольку шифрование Пэйе гомоморфно относительно умножения на незашифрованные константы и относительно сложения, то все вычисления можно выполнять в зашифрованном виде, а результат вычисления этой линейной формы будет представлен в зашифрованном виде.

Если $b_{k+1}, \ldots, b_n$ не все равны нулю, то система не имеет решений. В противном случае она имеет бесконечное множество решений. Переменным $x_{k+1}, \ldots, x_n$ можно присвоить произвольные значения, а переменные $x_1, \ldots, x_k$ выражаются через них.

4. Заключение

С точки зрения криптографических приложений свойство гомоморфности функции шифрования не всегда должно однозначно оцениваться как достоинство криптосистемы. Известны примеры, когда это свойство следует отнести к слабостям. Так, преобразование, обратное функции шифрования криптосистемы RSA, используется в схеме электронной подписи RSA для генерации подписей. Подпись сообщения m вычисляется по формуле $s = m^d \bmod N$, где d — секретная экспонента. Очевидно, что и это обратное преобразование гомоморфно относительно операции произведения сообщений. В результате, всегда имеется следующий способ подделки подписей. Если известны подписи s_1 и s_2 сообщений m_1 и m_2 соответственно, то подпись сообщения $m_1 \cdot m_2$ можно получить в виде $s_1 \cdot s_2$. На практике это не представляет никакой угрозы стойкости

схемы электронной подписи RSA, поскольку подписываются не сами сообщения, а значения хэш-функции от сообщений. Однако, гомоморфность функции генерации подписей накладывает на хэш-функцию дополнительное требование, которое, вообще говоря, не следует из стандартных определений криптографической хэш-функции.

Криптосистемы с гомоморфными функциями шифрования не могут обладать свойством неподатливости (non-malleability). Неформально свойство неподатливости означает, что по данной криптограмме $E(m_1)$ противник не может сгенерировать другую криптограмму $E(m_2)$ такую, что открытые тексты m_1 и m_2 связаны некоторым соотношением. Это понятие очевидным образом обобщается на случай, когда исходными данными для противника служат две или более криптограмм. Несколько более подробную информацию о неподатливой криптографии можно найти в работе [3].

Литература

[1] D. Bonech, Eu-Jin Goh, K. Nissim, Evaluating 2-DNF Formulas on Cipertexts, Proceedings of Theory of Cryptography Conference 2005. 31

[2] A.V. Shokurov, On measures of resistance of data encodings. Intitute for Sistem Programming Russian Acad. of Sci., Technical Report, February 2001. 32

[3] Варновский Н. П. Математическая криптография. Несколько этюдов. Материалы конференции "Московский университет и развитие криптографии в России". МГУ, 17–18 октября 2002 г., МЦНМО, 2003, 98–121 36

[4] Paillier P. Public-key cryptosystem based on composite degree residuosity classes, EUROCRYPT'99, Lecture Notes in Computer Science, 1592, 223–238 33

[5] Barak B., Goldreich O., Impagliazzo R., Rudich S., Sahai A.,.Vedhan S., Yang K. // On the (Im)possibility of obfuscating.programs. CRYPTO'01 - Advances in Cryptology, Lecture Notes in.Computer Science, v. 2139, 2001, p. 1-18. 29