N. E. Arykov, S. F. Krendelev

Developing a White-Box Public Key Cryptography and Its Application for Building a Secure Data Storage

Keywords: public key cryptography, hash functions, secure data storage

In the article existing cryptographic solutions of protecting confidential data were analyzed and the basic problem of a small number of secret parameters was formulated. This paper describes an algorithm that depends on a large number of secret parameters, based on one-to-one functions and two applications that allow to encrypt user passwords.

Н. Е. Арыков, С. Ф. Кренделев

РАЗРАБОТКА WHITE-BOX КРИПТОГРАФИЧЕСКОЙ СИСТЕМЫ И ЕЕ ПРИМЕНЕНИЕ ДЛЯ БЕЗОПАСНОГО ХРАНИЛИЩА ДАННЫХ[1]

Введение

С ростом вычислительных мощностей классические способы защиты хранилищ данных становятся с каждым годом все более уязвимыми. Популярные методы, такие как хеш-функции и криптографические протоколы с открытым ключом, например RSA или Elgamal, имеют ряд недостатков. Основные из них — малое количество используемых параметров, а также значительное увеличение скорости шифрования и дешифрования с ростом ключа [1]. За 2012 г. произошли утечки баз данных паролей огромных хранилищ данных, таких как LinkedIn, Yahoo и других. Процент расшифровки паролей велик, что показывает необходимость в создании криптографической системы, обладающей большей криптостойкостью, по сравнению с существующими методами, за счет увеличения количества используемых параметров и отсутствия свойства гомоморфности. Также описывается применение системы для шифрования данных в СУБД MySQL.

1. Анализ существующих решений

1.1. Хеш-функции

Главная задача хеш-функций — обеспечение целостности и конфиденциальности. Примером обеспечения конфиденциальности является хранение паролей пользователей. Например, в базах данных, файле /etc/shadow на unix-системах, а также в маршрутизаторах Cisco. В случае утечки базы данных паролей злоумышленник не должен иметь возможности восстановить оригинальный пароль. Для обеспечения целостности информации используются расширения хеш-функций, такие как HMAC, MAC.

Главной проблемой хеш-функций является использование одного секретного параметра — типа функции. При этом список популярных хеш-функций не является достаточно большим: MD5, SHA-1, SHA-2, RIPEMD-256, BCrypt.

1.2. Криптографические протоколы с открытым ключом

Хеш-функции могут быть заменены криптографическим протоколом с открытым ключом, например RSA. RSA является устойчивым к атаке коллизий, но приводит к новым уязвимостям. Секретный ключ RSA состоит только из двух параметров. Самыми распространенными криптосистемами с открытым ключом являются RSA и Elgamal, данные системы обладают свойством гомоморфности относительно умножения, которое позволяет производить атаку по выбранному шифротексту.

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

Опишем гомоморфность RSA. В качестве односторонней функции алгоритм RSA использует умножение простых чисел, основываясь на вычислительной сложности задачи факторизации. RSA обладает свойством гомоморфности относительно умножения. Пусть m_1, m_2 — различные открытые тексты, c_1, c_2 — соответствующие им шифротексты. Опишем процесс шифрования сообщения $m_1 * m_2$, $(m_1 * m_2)^e = m_1^e * m_2^e = c_1 * c_2 (\mathrm{mod}\, n)$, что эквивалентно $E(k, m_1 m_2) = E(k, m_1) * E(k, m_2)$. Из данного преобразования видно, что из двух открытых текстов можно получить третий путем объединения, шифротекстом для которого будет объединение шифротекстов c_1, c_2.

1.3. Анализ массовых утечек баз данных паролей

За 2012 г. было несколько случаев утечек баз данных паролей больших порталов, приведем таблицу со статистикой.

Сервер/Владелец	Год	Средняя длина пароля	Паролей проверено	Процент	Хеш-функция
Yahoo	2012	8	453000	—	Plain text
LinkedIn	2012	—	6500000	58	SHA1 (unsalted)
Last.fm	2012	—	17300000	95	MD5 (unsalted)
Formspring	2012	—	420000	—	sha-256 with random salt

Из приведенной таблицы видно, что процент расшифровки достаточно велик. В первую очередь это связано с малой длиной паролей пользователей. Во-вторых, скорость расчета используемых хеш-функций, таких как MD5 и SHA1, не устойчива к полному перебору на графических вычислительных устройствах (GPU). Высокая скорость перебора связана с малым количеством используемых параметров хеш-функций.

2. Конструирование криптографических хеш-функций

Обозначим через R конечное кольцо, $q = |R|$ — мощность множества R, $R^n = R \times R \times \ldots \times R$ — декартовое произведение. Рассмотрим произвольное отображение $H : R^n - R^m$, если $n > m$, то будем называть H хеш-функцией. Всякое отображение H полностью определяется набором функций $h_i = R^n \to R (i = 1,2,\ldots,m)$. Согласно [2], набор функций h_i называется ортогональным, если система уравнений $y_i = h_i(x_1, x_2, \ldots, x_n)$, где $i = 1 \ldots n$, имеет в точности q^{n-m} решений для всякого набора $(y_1, y_2, \ldots, y_3)$. В частности, если $n = m$, то ортогональность означает, что отображение H взаимно-однозначно. Если $m = 1$, то многочлен в системе уравнений единственный и по терминологии [2] называется перестановочным многочленом. Условие криптографической стойкости эквивалентно тому, что хеш-функция построена из ортогональной системы. Действительно, образом отображения является все множество R^m, так как система уравнений разрешима при любой левой части. В силу того, что для всякой левой части количество решений одинаково, коллизии распределены равномерно. Остается условие того, что при заданной левой части найти решение системы уравнений является сложной задачей.

В том случае, когда кольцо R является конечным полем F_q, где q — количество элементов в поле, имеет место два факта. Первый: любая подсистема ортогональной системы является ортогональной. Этот факт получается из того, что в случае конечных полей любую функцию можно представить как многочлен. Второй факт: для любой ортогональной системы $h_1, h_2, \ldots, h_3$, $m < n$, для всякого $0 < r \le n - m$, существуют многочлены $h_{m+1}, \ldots, h_{m+r}$, такие,

что система $h_1, h_2, \ldots, h_m, h_{m+1}, \ldots, h_{m+r}$ является ортогональной. Будем считать очевидным перенесение результатов для полей на кольца типа Z_p или их расширения. Отсюда следует, что если требуется построить хеш-функцию с необходимыми свойствами, то исходной точкой должно быть построение взаимно-однозначного отображения из множества R^n на себя.

Рассмотрим варианты построения криптографических хеш-функций. Для полей известные взаимно-однозначные отображения имеют вид $h(x) = x^k$ и $h(x) = ax + b$, причем любое другое взаимно-однозначное отображение получается суперпозицией такого типа отображений. Зафиксируем конечное поле F и набор таких отображений $h_1(x_1), h_2(x_2), \ldots, h_k(x_k)$. Отображение $H: F^k \to F^k$, определенное по формуле $y_i = h(x_i)$, $i=1,2,\ldots,k$, назовем диагональным. Отметим, что такого сорта преобразования в криптографии эквивалентны полиалфавитному шифрованию. Для стойкости шифрования диагональное отображение небезопасно, поскольку нет зависимости между элементами. Необходимо сделать дополнительное преобразование для обфускации. Построим новый набор отображений $g_1(x_1, x_2, \ldots, x_k), g_2(x_1, x_2, \ldots, x_k), \ldots, g_k(x_1, x_2, \ldots, x_k)$ по правилу $g_i(x_1, x_2, \ldots, x_k) = a_{i1}h_1(x_1) + a_{i2}h_2(x_2) + \ldots + a_{ik}h_k(x_k)$, $i=1,2,\ldots,k$, где коэффициенты a_{ij} определяют обратимую матрицу над полем.

Для реализации в рамках white-box-криптографии данные хеш-функции не являются подходящими, поскольку структура отображения довольно очевидна. Каждое отображение явно представляется в виде суммы слагаемых, зависящих от одного переменного. Для обфускации независимых переменных необходимо построить взаимно-однозначное отображение $x_i = g_i(z_1, z_2, \ldots, z_k)$, где $i = 1..m$. Затем сделать замену переменных в исходном отображении. В качестве отображения можно взять преобразование Кремоны или сети Фейстеля. В простейшем случае можно взять линейное преобразование, тогда можно рассматривать отображения из векторного пространства F^n в векторное пространство F^k, где $n > k$, причем значение n может быть произвольным.

3. Конструирование криптографического протокола с открытым ключом

Рассмотрим предлагаемую систему шифрования, зависящую от большого количества параметров, используя описанную теорию для криптографических хеш-функций. Пусть p — простое число, тогда кольцо вычетов по модулю p — Z_p является полем. Для шифрования необходимы только такие многочлены, которые являются взаимно-однозначными отображениями, — перестановочные многочлены.

Согласно теореме о перестановочных многочленах, S_p порождается полиномами cx, $x+1$, x^{p-2} [2]. Также одночлен x^n является перестановочным многочленом поля $F_q \Leftrightarrow$ НОД (n, $q-1$) =1, то есть числа n и $q-1$ взаимно простые. Обратным отображением к отображению x^k будет отображение x^d, где d — обратный элемент k в кольце Z_{p-1}. Обратимым отображением будет аффинное отображение вида $ax+b$, где $a \neq 0$. Не зная p, как и в RSA, найти обратный элемент можно только с помощью метода полного перебора.

Следовательно, перестановочный многочлен над полем может быть описан как различная суперпозиция отображений, таких, что: $P(x) = (ax+b)^e + d$, где $a \in Z_p$, $a \neq 0$, b, d — произвольные элементы из Z_p, e — обратимый элемент из $Z_{\varphi(p)} = Z_{p-1}$ ($\varphi(p)$— функция Эйлера). Такие отображения $P(x)$ назовем элементарными. Число всех взаимно-однозначных отображений можно легко посчитать — оно равно порядку симметрической группы S_p, то есть группе всех перестановок на множестве из p элементов — $p!$. Учитывая данные теоретические основы, предлагается следующая схема шифрования:

Выбираются простые числа $p_1 \cdots p_r$. Модулем является их произведение

$$n = \prod_{i=1}^{r} p_i .$$

- Для каждого числа p_i выбирается набор элементарных полиномов $f_1^{p_i}(x)...f_d^{p_i}(x)$ над Z_{p_i} и вычисляется полином $F_{p_i}(x) \equiv f_1^{p_i}(...(f_d^{p_i}(x)...)(\mathrm{mod}\, Z_{p_i})$. После раскрытия скобок в $F_{p_i}(x)$ получается взаимно-однозначный полином над полем Z_{p_i}.
- При помощи китайской теоремы об остатках и расширенного алгоритма Евклида вычисляется полином $F(x)$ над полем Z_n: $F(x) = F_{p_i}(x)(\mathrm{mod}\, p_i) i = 1...r$.

Открытый ключ состоит из взаимно-однозначного многочлена $F(x)$ и модуля n. Закрытый ключ состоит из простых чисел, элементарных полиномов и порядка раскрытия скобок. Длина публичного ключа в битах равна произведению длины модуля в битах, умноженному на количество ненулевых коэффициентов многочлена.

Процесс шифрования выглядит следующим образом. Пусть Боб обладает публичным ключом Алисы (F, n) и хочет передать ей секретное сообщение m по открытому каналу. Для шифрования Боб переводит сообщение m в числовое представление и вычисляет шифротекст $c = F(m)(\mathrm{mod}\, n)$, после чего передает его Алисе. Алиса хочет восстановить сообщение m из шифротекста c, используя свой приватный ключ. Для дешифрования используется следующий алгоритм:
- Алиса вычисляет числа $C_i \equiv c(\mathrm{mod}\, p_i)$, $i = 1...r$.
- Затем Алиса решает систему уравнений $F_{p_i}(x) \equiv C_i(\mathrm{mod}\, p_i)$, $i = 1...r$, используя информацию о построении отображений $F_{p_1}...F_{p_s}$. В результате Алиса получает систему $x \equiv m_i(\mathrm{mod}\, p_i)$ и, используя расширенный алгоритм Евклида, получает m.

4. Безопасность предлагаемой криптосистемы

Предлагаемая система не является гомоморфной, так как по построению взаимно-однозначного полинома мономы запрещены. Следовательно, система устойчива к атакам по подобранному и адаптивно подобранному шифротекстам. Наилучшее известное время для взлома данной криптосистемы — субэкспоненциальное [3], как black-box-задачи над полем простых чисел. Криптосистема основывается не только на сложности факторизации целых чисел, как RSA, но и на сложности разложения многочленов. Наилучшие алгоритмы факторизации работают за субэкспоненциальное время. Главной отличительной особенностью алгоритма является то, что секретных параметров намного больше, чем в криптосистемах RSA и Эль-Гамаля. В экспериментах, которые описаны в работе [4], доказано, что длина цикла значений многочлена $F(x)$ над кольцом Z_n больше, чем многочленов $f_1^{p_i}(x)...f_d^{p_i}(x)$ над полем Z_{p_i}, $i = 1...r$. Таким образом, данная система устойчива к атаке на циклы. В дальнейшем криптосистему планируется сделать вероятностной, что позволит принципиально усилить шифрование.

5. Применение

В качестве применения описанной криптосистемы приводятся два разработанных приложения для шифрования секретной информации пользователей. Программа «Encryptor creator» создает новую программу «Encryptor» с использованием кодогенерации. На вход данной программе пользователь подает:
- множество простых чисел $p_1, .., p_n$;
- множество элементарных полиномов $f_i^{p_j}$. Для задания полиномов необходимо указать степень многочлена на отрезке [a, b], после чего выбираются коэффициенты из своего доверительного интервала;
- порядок раскрытия скобок.

Результатом работы данной программы является исполняемый файл «Encryptor». Он содержит биективный многочлен и модуль. Данное приложение получает на вход секретные данные, например пароль, и шифрует его. Владелец хранилища паролей определяет набор параметров вручную или с

помощью генератора случайных чисел, а затем получает приложение, которое может использоваться для шифрования паролей пользователей.

Вторая реализация — это плагин для СУБД MySQL и phpMyAdmin. Владелец устанавливает параметры для каждой базы данных, содержащей пароли пользователей. Каждая база данных связывается с биективным полиномом и модулем. Шифрование реализовано в виде триггера, позволяющего шифровать кортежи и отношения в базе данных.

СПИСОК ЛИТЕРАТУРЫ:

1. *Alese B. K., Philemon E. D.* Comparative Analysis of Public-Key Encryption Schemes // International Journal of Engineering and Technology. 2012. September. Vol. 2. № 9.
2. *Lidl R., Niederreiter H.* Finite fields. Cambridge University Press, 1984.
3. *Boneh D., Lipton J. R.* Algorithms for black-box fields and their application to cryptography // 16th Annual International Cryptology Conference. 1996. T. 1109. P. 283–297.
4. *Krendelev S. F., Spitsyna E. O.* Number 667. Security analysis of RSA. Variant of public-key cryptography. Journal "Highly available systems", 2011. P. 34–38.

REFERENCES:

1. *Alese, B. K., Philemon E. D.* Comparative Analysis of Public-Key Encryption Schemes, International Journal of Engineering and Technology Vol. 2 No. 9, September, 2012.
2. *Lidl R., Niederreiter H.* "Finite fields" Cambridge University Press, 1984
3. *Boneh D., Lipton J. R.* Algorithms for black-box fields and their application to cryptography // 16th Annual International Cryptology Conference, 1996. — T. 1109. C. 283–297.
4. *Krendelev S. F., Spitsyna E. O.* (2011) Number 667. Security analysis of RSA. Variant of public-key cryptography.