КРИПТОЛОГІЯ / CRYPTOLOGY

DOI: 10.18372/2225-5036.22.11097

РАСПРЕДЕЛЕНИЕ ДАННЫХ НА ОСНОВЕ КОДОВ, ИСПРАВЛЯЮЩИХ ОШИБКИ

Артак Хемчян

Национальный политехнический университет Армении, Армения

ХЕМЧЯН Артак Араратович

Год и место рождения: 1990 год, Ереван, Армения.
Образование: Национальный политехнический университет Армении, 2014 г.
Должность: аспирант 2 года.
Научный интерес: информационная безопасность, облачные системы, криптография, стеганография, коды исправляющие ошибки.
Публикации: около 10 публикаций по теме распределение данных и информационной безопасности.
E-mail: artak-khemchyan@mail.ru

Аннотация. *Протоколы разделения секрета применяются для распределенного хранения информации. Чаще всего такой информацией оказываются секретные ключи или пароли какого-либо абонента. Эти протоколы призваны решить проблему хранения информации так, чтобы те группы людей, которым позволено знать секрет, могли бы его восстановить, а те группы, которым секрет знать не позволено, восстановить его не смогли даже путем перебора. Более популярной пороговой схемой распределения секрета является схема распределения Шамира. Идея, на которой основана схема Шамира, заключается в том, что для интерполяции многочлена степени* $k-1$ *требуется* k *точек. Схема распределение Шамира изначально был предназначен для распределения секретных ключей. При больших объемах данных процесс распределения и восстановления данных проходит очень медленно. Таким образом, необходимо иметь более быструю пороговую схему распределения данных. Так как схемы распределение секрета основаны на избыточность, и избыточность обеспечивают коды исправляющие ошибки, возникает идея построить схему распределение данных на основе кодов. В статье описано новая система распределения секрета на основе кодов, исправляющих ошибки. Целью исследования является разработка высокопроизводительной системы, которая будет использовать разные коды, исправляющие ошибки.*

Ключевые слова*: распределение секрета, пороговая схема, схема Шамира, коды, исправляющие ошибки, код Хэмминга.*

Введение

В настоящее время значительная часть глобального обмена данными происходит через открытые сетевые соединения. Большое количество данных доступно практически всем. Необходимо гарантировать информационную безопасность и защиту [1]. Вместе с применением шифрования, секретные ключи должны храниться в защищенном хранилище. Секретные ключи непосредственно не должны быть доступны где-либо в системе без заранее обеспеченной физической защиты. Увы, это условие практически невыполнимо.

В целях обеспечения безопасности ключей применяются пороговые схемы распределения секрета: ключ делится на несколько частей и хранится в разных местах.

Распределение секрета

Протоколы распределения секрета применяются для распределенного хранения информации. Чаще всего такой информацией оказываются секретные ключи или пароли какого-либо абонента. Например, главный бухгалтер предприятия держит секретную рабочую информацию зашифрованной, а ключ длиной 64 бита хранит в надежном месте, известном только ему. Но что случится, если главный бухгалтер уволится или тайник с ключом сгорит? Тогда ключ будет утерян, а дешифрование в современных криптосистемах может занять миллионы лет! Можно выдать по копии ключа заместителю главного бухгалтера и директору предприятия. Но если заместитель захочет занять место главного бухгалтера, воспользуется своей копией ключа и подменит важную информацию? Или продаст ее конку-

рентам? Можно предложить следующий выход: разделить ключ на четыре части по 16 бит и выдать одну часть генеральному директору, другую – его заместителю, третью – заместителю главного бухгалтера, а четвертую – супругу (супруге) главного бухгалтера. Но что, если заместители договорятся сместить своих начальников и воспользуются своими частями ключа? Тогда для того, чтобы восстановить ключ, злоумышленникам потребуется подобрать всего лишь 32 бита, что потребует всего $2^{32} \approx 4,3 \times 10^9$ операций вместо $2^{64} \approx 18,5 \times 10^{18}$ при подборе 64 битов. Злоумышленники смогут восстановить ключ за вполне обозримое время.

Разумным будет разделить этот 64-битовый ключ K так, чтобы каждому досталось по 64 бита. Как это сделать? Генеральному директору, и заместителям можно выдать по случайной 64-битовой строке S_1, S_2, S_3 соответственно, а супругу(е) главного бухгалтера – строку: $S_4 = K - S_1 - S_2 - S_3 (mod\, 2^{64})$. Тогда каждый из них будет обладать случайной строкой бит, по которой ключ можно восстановить только перебором 64-битового числа. Даже соединив три любых части, нельзя получить никакой информации о ключе, и нельзя уменьшить количество перебираемых битов. Но, при соединении всех четырех частей ключ вычисляется однозначно:

$$S_1 + S_2 + S_3 + S_4 \equiv K (mod\, 2^{64}) .$$

Выше описана простейшая схема распределения секрета с одной разрешенной группой участников, состоящей из 4-х абонентов.

Протоколы распределения секрета призваны решить проблему хранения информации так, чтобы те группы людей, которым позволено знать секрет, могли бы его восстановить, а те группы, которым секрет знать не позволено, восстановить его не смогли даже путем перебора.

В протоколе распределения секрета имеются n участников (абонентов) $P_1, P_2, \ldots, P_n$ и один выделенный участник D , называемый дилером (раздающим). Пусть через $P = \{P_1, P_2, \ldots, P_n\}$ обозначено множество всех абонентов. Введем следующую терминологию: группа доступа (разрешенная группа) – непустое подмножество A участников множества P , которые, собравшись вместе, имеют право восстановить секрет; структура доступа Г будем называть непустое множество всех групп доступа.

Далее будем полагать, что любой участник $P_1, P_2, \ldots, P_n$ входит хотя бы в одну группу доступа, иначе его присутствие бессмысленно. Также считаем, что Г замкнуто, то есть если $A \subset B \subset P u A \subset \Gamma$, то $B \subset \Gamma$. Действительно, если абоненты $P_1, P_2, \ldots, P_k$ могут совместно восстановить секрет, то, если к ним присоединятся дополнительные участники P_{k+1}, P_{k+2},..., то получившаяся группа тем более сможет восстановить секрет.

Наиболее распространены пороговые схемы распределения секрета.

Определение 1: (k, n) – пороговой схемой распределения секрета $(k \le n)$ называется такая схема, в которой секрет распределяется между n участниками, причем разрешенной группой является любая группа из не менее, чем k участников. Более популярной пороговой схемой распределения секрета является схема распределения Шамира [4].

Рассмотрим схему Шамира и его особенности.

Схема Шамира

Идея, на которой основана схема Шамира, заключается в том, что для интерполяции многочлена степени $k-1$ требуется k точек. Если известно меньшее количество точек, то интерполяция будет невозможной. Обозначим: p – большое простое число (больше любого секрета M , который предполагается разделять в этой схеме). Тогда $M \in Z_p : n$ – число долей секрета; k – минимальный размер разрешенной группы.

Работа алгоритма можно разделить на 3 этапа.

Подготовительный этап

Дилер выбирает случайным образом коэффициенты $S_1, S_2, \ldots, S_{k-1} \in Z_p$ и составляет секретный многочлен

$$S(x) = S_{k-1} X^{k-1} + S_{k-2} X^{k-2} + \ldots + S_1 X + M \; mod \; p ,$$

где M – разделяемый секрет, а остальные коэффициенты – произвольные элементы поля (коэффициенты многочлена дилер хранит в тайне). Очевидно, что $S(0) = M$. Далее дилер выбирает n различных несекретных ненулевых элементов $r_1, r_2, \ldots, r_n$ из Z_p , каждый из которых ставит в соответствие одному участнику схемы.

Распределение секрета

Дилер вычисляет значения многочлена $c_1 = S(r_1), c_2 = S(r_2), \ldots, c_n = S(r_n)$. Доля каждого пользователя A_i – это пара чисел $(r_i, c_i), i = 1, 2, ..., n$. Доли раздаются участникам схемы.

Восстановление секрета

Чтобы восстановить секрет M , надо воспользоваться интерполяционной формулой Лагранжа: если нужно построить многочлен $S(x)$ степени $(k-1)$, который при $x_1, x_2, \ldots, x_k$ принимает соответственно значения $y_1, y, \ldots, y_k$, то этим многочленом будет: $S(x) = \sum_{j=0}^{k-1} y_i \prod_{j \ne i} \frac{x - x_j}{x_i - x_j}$. k как в схеме разделения секрета многочлен положено выбрать так, чтобы $S(0)$ =M, то из формулы Лагранжа следует:

$$M = \sum_{i=0}^{k=1} c_i S_i , \text{ где } S = \prod_{i \ne j} \frac{r_j}{r_j - r_i} .$$

Из описанного выше, становится ясно, что для больших значений порога, вычисление становится медленно.

Распределение данных. Первоначально пороговые схемы разделения секрета были предназначены для распределения секретных ключей. Это означает, что безопасность данных обеспечивается следующим образом: конфиденциальная информа-

ция шифруется; ключ шифрования распределяется между сторонами по (n,k) пороговой схемой.

В этом случае, мы гарантируем конфиденциальность данных, и только авторизованная группа участников может восстановить ключ шифрования, а затем расшифровать зашифрованную информацию. Но это не решит проблемы целостности и доступности данных. Представим себе ситуацию, когда повреждена зашифрованная информация. В этом случае, стороны восстановив ключ шифрования, тем не менее, не будут иметь возможность полностью дешифровать секретные данные. Оказывается, что такой вариант распределения данных не гарантирует целостность данных. Теперь представим себе ситуацию, когда секретная информация на некоторое время не доступно (сервер недоступен, файл был удален, и т.д.). В этом случае тоже, стороны восстановив ключ шифрования, тем не менее, не будут иметь возможность дешифровать секретные данные.

Для решения этих проблем, предлагается распределить данные целиком, (n,k) пороговой схемой. В этом случае обеспечивается конфиденциальность, целостность и доступность данных.

Как уже упоминалось ранее схема распределения Шамира изначально была предназначена для распределения секретных ключей. При больших объемах данных процесс распределения и восстановления данных очень медленно. Таким образом, необходимо иметь более быструю пороговую схему распределения данных. Предлагается построить новую пороговую схему распределения секрета, основанную на кодах, исправляющие ошибки.

Метод разделения. В этой системе используются следующие коды: Хэмминга [2, 3], Рида-Мюллера [2, 3]; Рида-Соломона [2], БЧХ [2, 4].

Ниже описывается принцип работы метода распределения секрета на примере кода Хэмминга (7, 4, 1). После кодирования 4-битной исходной информации получим 7-битную кодированную информацию, которую назовем кодовым словом. Это позволит нам корректировать один ошибочный бит. Каждый 4 бит секретного файла (который надо распределить) кодируется кодом Хэмминг (7, 4, 1), и получаются соответствующие 7-битовые кодовые слова. Эти кодовые слова представим в виде матрицы, как показано на рис. 1. Понятно, что этот файл после кодирования кодом Хэмминга (7, 4, 1) нам надо распределить.

В этом примере k – длина кодового слова (в данном случае $k=7$), а q зависит от объема исходного файла. Каждая строка массива представляет собой кодовое слово кода Хэмминга (7, 4, 1), это означает, что любая одна ошибка может быть исправлена в этих 7 битах. На основе этой характеристики мы можем распределить по столбцам. Если рассматривать каждый столбец в качестве отдельного компонента, то очевидно, что мы можем восстановить оригинальный файл в случае повреждения или потери любого компонента. Это приводит к $(6,7)$ пороговой схеме.

Группировка. Чтобы увеличить безопасность и иметь компоненты, равные по объему распределяемого файла, мы должны применить группировку по определенному методу. От каждой части 4 столбцов должны быть взяты данные (с определенным методом), для удовлетворения всех требований. Например, в случае группировки согласно табл. 1 мы получим $(2,4)$ пороговую схему.

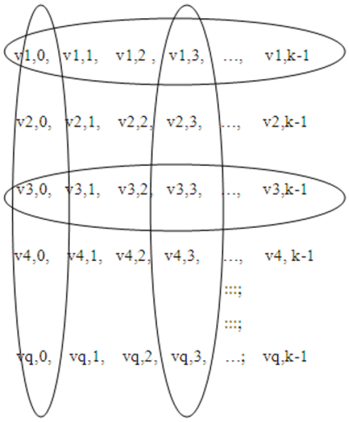

Рис. 1. Структура кодированного файла

Пороговая схема (2, 4) Таблица 1

Часть 1	3	6	2	7
Часть 2	1	6	7	5
Часть 3	5	4	2	6
Часть 4	7	5	3	4

Каждая строка представляет собой отдельную часть (компонент). Как мы видим, часть 1 содержит количество столбцов, взятых с определенным методом. Табл. 1 показывает, что после слияния любых трех частей, у нас будет 6 частей, и один будет отсутствовать. При использовании алгоритма декодирования кода Хэмминга (7, 4, 1) мы можем восстановить одну недостающий (испорченную) часть. Оказывается, распределение секрета в соответствии с таблицей 1 приводит к пороговой схеме $(2,4)$. Секретный файл разделен на четыре части так, что любые три части достаточны для восстановления. Табл. 2 представляет $(2,3)$ пороговую схему.

Пороговая схема (2,3) Таблица 2

Часть 1	5	2	6	4
Часть 2	4	7	3	5
Часть 3	6	2	7	3

Сравнение. На рис. 2 показано график распределения файла (объемом 10Мб) методом Шамира ($n=5,\ k=2,3,4,5$).

Экспериментальные данные были получены на компьютере со следующими параметрами: CPU – Intel Core i3 2.00 Ghz; RAM – 2GB. Сравнивая графики из рис. 2 и 3 видим, что распределение кодом Хемминга намного быстрее.

Тот же метод разделения секрета применим для других кодов, исправляющих ошибки. В дальнейшем планируется внедрять другие коды тоже, сравнивая все коды по быстродействию. Система станет автоматически выбирать тот код, который в данном случае будет самым быстрым. Эти действия будут невидимы для пользователя. Продолжая исследования, станем проводить разные сравнитель-

ные тесты по разным параметрам с методами Шамира и с другими методами разделения секрета.

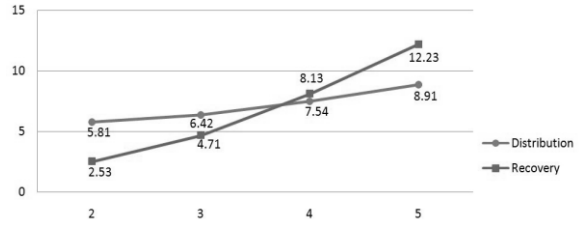

Рис. 2. График распределения и восстановления файла методом Шамира

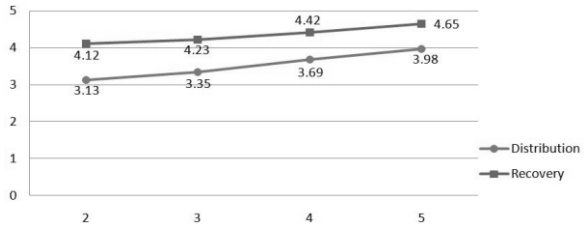

Рис. 3. График распределения и восстановления файла кодом Хемминга

Выводы

Таким образом, было показано, что полное распределение данных обеспечивает не только конфиденциальность, но и целостность, и доступность данных. Для этой цели, создана новая схема распределение секрета на основе кодов, исправляющих ошибки. Этот метод намного быстрее, так как использует коды, исправляющие ошибки, а эти коды используют логические операции. Продолжая исследование, будем тестировать этот метод с методом Шамира по разным параметрам.

Литература

[1] Шнайер Б. Прикладная криптография / Шнайер Б. // М. : Издательство ТРИУМФ, 2003. – С. 17-29.

[2] Peterson W.W. Error-Correcting Codes / Peterson W.W., Weldon E.J. // The Massachusetts Institute of Technology, Second Edition – 1972 – P. 301-350.

[3] Assmus E.F. Designs and Their codes / Assmus E.F., Key J.D. // Cambridge University Press – 1992. – P. 264-270.

[4] Bose R. On A Class of Error Correcting Binary Group Codes / Bose R., Ray-Chaudhuri D.K., // Information and Control V.3 – 1960. – P. 68-79.

УДК 004.056.3 (045)

Хемчян А. Обмін даними на основі кодів корегування помилок

Анотація. *Секретні протоколи обміну використовуються для розподіленого зберігання інформації. Найчастіше така інформація – це секретні ключі або паролі будь-якого користувача. Секретні протоколи обміну призначені для вирішення проблеми зберігання інформації так, щоб ті групи людей, яким дозволено знати секрет, могли відновити таємницю, а також ті групи, які не можуть знати секрет, не в змозі відновити секрет навіть за допомогою перебору. Більш популярна порогова схема поділу секрету є схема розподілу Шаміра. Ідея схеми Шаміра – для інтерполяційного полінома ступеня к-1, до точок потрібно к точок. Схема поділу Шаміра була спочатку розроблена для розподілу секретних ключів. Для великих обсягів даних процес обміну і відновлення відбувається дуже повільно. Таким чином, необхідно мати більш швидку порогову схеми розподілу даних. Оскільки секретна схема заснована на надлишковості інформації, і надлишковість забезпечують кодами, які виправляють помилки, постає ідея побудувати схему розподілу на основі цих кодів. У статті описується нова система таємного обміну на основі кодів, що виправляють помилки. Мета дослідження полягає в тому, щоб розробити високоефективну систему, яка буде використовувати різні коди корекції помилок.*

Ключові слова: *поділ секрету, порогова схема, схема Шаміра, коди виправлення помилок, код Хеммінга.*

Khemchyan A. Data sharing based on error-correcting codes

Abstract. *Secret sharing protocols are used for distributed storing of information. Most often such information are secret keys or passwords of any user. Secret sharing protocols are designed to solve the problem of storing information so that those groups of people, who are allowed to know the secret, can restore the secret, and those groups that are not allowed to know the secret, are not able to recover even by brute-force. More popular threshold secret sharing scheme is Shamir's distribution scheme. The idea of Shamir's scheme is the fact that for the interpolation polynomial of degree k-1, k points are required. Shamir's sharing scheme was originally designed for distribution of secret keys. For large amounts of data, sharing and restore processes are very slow. Thus, it is necessary to have a faster threshold secret sharing scheme. Since the secret scheme is based on the redundant information, and error-correcting codes provide redundancy, the idea was came to build a distribution scheme based on these codes. This article describes a new system of secret sharing on the basis of error-correcting codes. The purpose of the study is to develop a high-performance system which will use a different error-correcting code.*

Key words*: secret sharing, threshold scheme, Shamir scheme, error-correcting codes, Hamming code.*

Отримано 4 *жовтня 2016 року,* затверджено редколегією 26 *жовтня 2016 року*