УДК 681.3.06

Г. З. ХАЛИМОВ, канд. техн. наук

КАСКАДНОЕ УНИВЕРСАЛЬНОЕ ХЕШИРОВАНИЕ ПО РАЦИОНАЛЬНЫМ ФУНКЦИЯМ АЛГЕБРАИЧЕСКИХКРИВЫХ

Универсальные семейства хеш-функций являются объектами полевых структур, характеризуются прозрачными комбинаторными свойствами и имеют доказуемую секретность. Применение для аутентификации каскадного хеширования позволяет получить наименьшую вероятность коллизии, сложность вычислений хешей для фиксированного объёма хешируемых данных и размерности конечного поля. Основные результаты представлены в работах [1 – 5]. В работе [1] показано, что вероятность коллизии можно эффективно уменьшить путем хеширования сообщений по нескольким независимо выбранным хеш-функциям. Стинсон рассмотрел композиционное хеширование со снятием ограничения на размер ключевых данных для строго универсального хеширования [2]. Букетное хеширование, допускающее параллельные алгоритмы вычислений и снимающее ограничение на сложность, предложено Рогавеем [3]. Оценки вероятности коллизии для каскадного универсальное хеширование со связкой хеша предыдущего каскада с текстом последующего каскада получены в [4]. Свойства многократного универсального хеширования по проективной прямой и кривой Эрмита представлены в [5].

В данной работе предлагается метод каскадного универсального хеширования по рациональным функциям алгебраических кривых на основе произведения функциональных полей. С этой целью в разд. 1 рассмотрено каскадное универсальное хеширование по рациональным функциям алгебраических кривых со связкой хеша предыдущего каскада с текстом последующего. В разд. 2 представлено многократное универсальное хеширование на основе произведения функциональных полей алгебраических кривых.

1. Каскадное универсальное хеширование со связкой хеша с текстом

Для универсального хеширования по рациональным функциям алгебраических кривых определение каскадного хеширования со связкой по хешу и тексту имеет следующее определение.

Определение 1 [4]. Пусть F_q – конечное поле, M – сообщение и $M = M_1 \| M_2$. Каскадное универсальное хеширование по рациональным функциям алгебраических кривых определяется выражением

$$Ch(M) = AGh_2\left(AGh_1(M_1) \| M_2\right), \tag{1}$$

где AGh_1, AGh_2 – универсальные схемы хеширования по рациональным функциям алгебраических кривых, $Ch(M)$ определяет универсальное семейство хеш функций $\varepsilon - AU$, где $\varepsilon = \max(\varepsilon_1, \varepsilon_2) + 1/\left|H^2\right|$, $\varepsilon_1, \varepsilon_2$ – соответственно вероятности коллизий для AGh_1 и AGh_2 хеширования.

Коллизионные свойства каскадной конструкции определяются утверждениями 1.2.

Утверждение 1 [4]. Если H_1 есть $\varepsilon_1 - U$ универсальный класс и H_2 есть $\varepsilon_2 - U$, тогда $H = H_1 H_2$ есть $\varepsilon - U$, где $\varepsilon = \max(\varepsilon_1, \varepsilon_2) + 1/\left|H^2\right|$.

Утверждение 2 [4]. Пусть H_1 и H_2 соответственно $\varepsilon_1 - U$ и $\varepsilon_2 - U$ универсальные классы хеш функций. Каскадная конструкция $H_1 H_2$ имеет наименьшую вероятность коллизии, если $\varepsilon_1 = \varepsilon_2$.

Замечание 1.

1. Каскадное хеширование определяется тем, что хеш предыдущего каскада связывается с текстом следующего каскада через конкатенацию.

2. Каскадное хеширование $Ch(M)$ при фиксированном поле вычислений предполагает разбиение данных на блоки приблизительно равной длины. Для двух каскадного универсального хеширования по проективной прямой вероятность коллизии уменьшится только в два раза, так как вероятность коллизии зависит пропорционально от значения длины данных. Для хеширования по максимальным кривым эта зависимость определяется как корень квадратный от длины сообщения и вероятность коллизии уменьшится только в $\sqrt{2}$ раз.

3. Вероятность коллизии каскадного хеширования имеет ограничение по наименьшему полю вычисления хеша одного из каскадов.

4. Размер ключевых данных увеличивается пропорционально числу каскадов, с учетом поля вычисления и универсального хеширования каскада.

5. Каскадное хеширование позволяет эффективно увеличить общую длину хешируемых данных и зафиксировать вероятность коллизии на уровне хеша первого каскада, если на втором и последующих каскадах увеличить поле вычислений. Примером является алгоритм хеширования UMAC(2000), каскадная схема применяется с подъёмом поля вычисления, сначала 32 бита, затем 64 бита и 128 бит.

Рассмотрим свойства каскадного хеширования без увеличения размера поля вычисления на примере двух каскадного хеширования по функциональному полю кривой Эрмита и проективной прямой.

Определение 2 [4]. Пусть F_q, $q = p^2$ – расширенное конечное поле, M – сообщение и $M = M_1 \| M_2$. Алгоритм вычисления хеш кода в каскадной конструкции определяется выражением

$$Ch(M) = Hh_q(PSh_q(M_1) \| M_2),$$

где Hh_q, PSh_q – универсальные схемы хеширования по кривой Эрмита и проективной прямой.

Свойства двух каскадного хеширования по кривой Эрмита и проективной прямой рассмотрены в [4].

Утверждение 3 [4]. Пусть F_q, $q = p^2$ – расширенное конечное поле, $M = M_1 \| M_2$, $M_1| \le \sqrt{q} + 1$, $|M_2| \le q\sqrt{q}$, $0 < k \le q\sqrt{q} + \sqrt{q}$. Тогда $Ch_q(M)$ определяет универсальное семейство хеш функций $\varepsilon - U\left(q^2\sqrt{q}, q^k, q\right)$, $\varepsilon = \max(\varepsilon_{PS}, \varepsilon_H) + 1/\left|q\sqrt{q}\right|$, $\varepsilon_{PS}, \varepsilon_H$ – соответственно, вероятности коллизий для PSh_q и Hh_q хеширования.

Замечание 2.

1. Для $Ch_q(M) = Hh_q(PSh_q(M_1) \| M_2)$ хеширования является справедливым утверждение 2. Вероятность коллизии будет минимальной, если $\varepsilon_H = \varepsilon_{PS}$.

2. Применение во внутреннем каскаде PSh_q хеширования дает преимущество в скорости вычислений, так как PSh_q хеширование определено над простым полем. Практический алгоритм вычислений может учитывать эту особенность. Для малых длин данных используется только PSh_q хеширование и при превышении длины подключается Hh_q хеширование.

3. Ключевое пространство увеличивается в q раз, вероятность коллизии фиксируется на уровне $\varepsilon_H = \varepsilon_{PS} = \sqrt{2k'/q}$, где k' определяется уравнением $k'^2 + k' = k$, k – число слов сообщения.

4. Каскадирование $Hh_q(PSh_q(M_1)\|M_2)$ приводит к небольшому увеличению размера хешируемых данных $k+\sqrt{k}$. Наибольшее увеличение хешируемых данных в два раза достигается, если на первом и втором каскадах используется одинаковая функция хеширования.

2. Каскадное хеширование на основе произведения функциональных полей

Дальнейшим развитием каскадного хеширования является хеширование на основе произведения функциональных полей.

Определение 3. Пусть F_q – конечное поле, M – сообщение и $M = M_1 \| M_2 \| ... \| M_t$. Алгоритм вычисления хеш кода в каскадной конструкции определяется выражением

$$Ch_t(M) = AGh_2\left(AGh_1(M_1) \| AGh_1(M_2) \| ... \| AGh_1(M_t)\right), \quad (2)$$

где AGh_1, AGh_2 – универсальные схемы хеширования по алгебраическим кривым, $Ch_t(M)$ определяет универсальное семейство хеш функций $\varepsilon - AU$, где $\varepsilon = \max(\varepsilon_1, \varepsilon_2) + 1/|H^2|$, $\varepsilon_1, \varepsilon_2$ – соответственно, вероятности коллизий для AGh_1 и AGh_2 хеширования.

Замечание 3.

1. Коллизионные свойства каскадного хеширования следуют из утверждения 1.

2. Каскадное хеширование $Ch_t(M)$ при фиксированном поле вычислений предполагает разбиение данных на t блоков равной длины. Для первого каскада вероятность коллизии определяется размером блока данных, для второго – значением t.

3. Вероятность коллизии каскадного хеширования имеет ограничение по наименьшему полю вычисления хеша одного из каскадов.

4. Размер ключевых данных определяется произведением пространства ключей первого и второго каскадов, с учетом поля вычисления и универсального хеширования каскада.

5. Каскадное хеширование $Ch_t(M)$ позволяет эффективно увеличить общую длину хешируемых данных и зафиксировать вероятность коллизии на уровне хеша одного из каскадов.

6. Двух каскадная конструкция $Ch_t(M)$ легко распространяется на многокаскадную $l - Ch_t(M)$, где l – число вложенных каскадов универсальных схем хеширования по алгебраическим кривым AGh_1, AGh_2,..., AGh_l.

Рассмотрим свойства каскадного хеширования $Ch_t(M)$ на примере двух каскадного хеширования по функциональному полю кривой Эрмита и проективной прямой.

Определение 4. Пусть F_q, $q = p^2$ – расширенное конечное поле, M – сообщение и $M = M_1 \| M_2 \| ... \| M_t$. Алгоритм вычисления хеш кода в каскадной конструкции определяется выражением

$$Ch_t(M) = Hh_q(PSh_q(M_1) \| PSh_q(M_2) \| ... \| PSh_q(M_t)). \quad (3)$$

где Hh_q, PSh_q – универсальные схемы хеширования по кривой Эрмита и проективной прямой.

Утверждение 4. Пусть F_q, $q = p^2$ – расширенное конечное поле, M – сообщение и $M = M_1 \| M_2 \| ... \| M_t$. $Ch_t(M)$ – хеширование (3). Тогда $Ch_t(M)$ определяет универсальное семейство хеш функций $\varepsilon - U\left(q^2\sqrt{q}, q^k, q\right)$, $\varepsilon = (2k)^{1/3}/q$, $0 < k \le q\sqrt{q}/2$.

Доказательство. Хеширование на каждом каскаде является универсальным и каскадная хеш функция также является универсальной. Пространство ключей определяется произведе-

нием числа ключей первого и второго каскадов и равно $q^2\sqrt{q}$. Пусть k число слов данных и k' – размер блока данных, $t = k/k'$. Наименьшая вероятность коллизии в силу утверждения 2 реализуется в случае $\varepsilon_H = \varepsilon_{PS}$. Подставим в выражения для ε_H и ε_{PS} значения k' и $t = k/k'$, получим $k' = (2k)^{1/3}$ и оценку для вероятности $\varepsilon = (2k)^{1/3}/q$. Так как ε имеет оценку сверху $\varepsilon \le 1/\sqrt{q}$, для $k' \le g$ получим $k \le q\sqrt{q}/2$, где g – род кривой Эрмита.

Замечание 4.

1. Каскадное хеширование по кривой Эрмита и проективной прямой $PSh_q - Hh_q$ имеет вероятность коллизии $\varepsilon = (2k)^{1/3}/q$, что совпадает с хешированием по кривой Сузуки $\varepsilon = 3k^{1/3}/q$ (см. [6]). Затраты по ключу несколько больше – $q^2\sqrt{q}$, в отличие от q^2 – для кривой Сузуки.

2. Вычисления в каскадном хешировании существенно проще. Вычисления по PSh_q каскаду выполняются на одной рациональной функции со сложностью $\sim k'$ и по Hh_q каскаду на двух рациональных функциях – со сложностью $\sim t + \sqrt{t}$. С учетом $t = k/k'$ и $k' = (2k)^{1/3}$ получим оценку для числа вычислений $k + t + \sqrt{t} = k + k^{2/3}/2^{1/3} + k^{1/3}/2^{1/6}$, что меньше почти в два раза числа вычислений по кривой Сузуки на четырёх рациональных функциях – $2k + 1.04k^{2/3} + 2\sqrt[3]{3}k^{1/3}$.

Возможны комбинации других универсальных хеш функций в двух каскадной схеме хеширования. Свойства каскадирования с кривой Ферма представлены утверждением 5.

Утверждение 5. Пусть F_q – конечное поле, M – сообщение и $M = M_1 \| M_2 \| \dots \| M_t$, $Ch_t(M)$ – хеширование вида (1), где $AGh_1 = PSh_q$, $AGh_2 = Fh_q$ – универсальные схемы хеширования по проективной прямой и кривой Ферма $x^{(q-1)/3} + y^{(q-1)/3} + 1 = 0$ соответственно. Тогда $Ch_t(M)$ определяет универсальное семейство хеш функций $\varepsilon - U\left(q^3, q^k, q\right)$, $\varepsilon = (9k/2)^{1/3}/q$, $0 < k \le q\sqrt{q}$.

Замечание 5. Доказательство подобно доказательству утверждения 4. Результаты каскадного хеширования $PSh_q - Fh_q$ совпадают по вероятности коллизии с хешированием $PSh_q - Hh_q$, но требуют больший размер ключа, соответственно q^3 и $q^2\sqrt{q}$.

Оценки многокаскадного универсального хеширования $l - Ch_t(M)$ по рациональным функциям алгебраических кривых представлены утверждениями 6-8.

Утверждение 6. Пусть F_q – конечное поле, $M = M_1 \| M_2 \| \dots \| M_t$ и $l - Ch_t(M)$ – l-каскадное универсальное хеширование по проективной прямой. Тогда $l - Ch_t(M)$ определяет универсальное семейство хеш функций $\varepsilon - U\left(q^l, q^k, q\right)$, $\varepsilon = k^{1/l}/q$, $0 < k \le q^l$, со сложностью вычислений $k + k^{l-1/l} + k^{l-2/l} - \dots + k^{1/l}$.

Доказательство. Хеширование на каждом каскаде является универсальным и каскадная хеш функция, также является универсальной. Пространство ключей определяется произведением числа ключей всех каскадов и равно q^l. Наименьшая вероятность коллизии в силу утверждения 2 реализуется, если на каждом каскаде значение вероятности коллизии является наименьшим. Это достигается, если размер данных хеширования k' на каждом каскаде является наименьшим $k' = k^{1/l}$. Подставим в выражения для ε_{PS} значения k' получим

$\varepsilon = k^{1/l}/q$ и оценку для $0 < k \le q^l$. Оценка сложности вычислений определяется тем, что на каждом каскаде число вычислений уменьшается в $k' = k^{1/l}$. Суммирование по всем каскадам дает результирующее выражение $k + k^{l-1/l} + k^{l-2/l} + ... + k^{1/l}$.

Утверждение 7. Пусть F_q, $q = p^2$ – расширенное конечное поле, M – сообщение и $M = M_1 \| M_2 \| ... \| M_t$, $l - Ch_t(M)$ – l-каскадное универсальное хеширование по кривой Эрмита. Тогда $l - Ch_t(M)$ определяет универсальное семейство хеш функций $\varepsilon - U\left(q^{l+l/2}, q^k, q\right)$, $\varepsilon = \sqrt{2k^{1/l}}/q$, $0 < k \le q^l$, со сложностью вычислений $k + k^{(2l-1)/2l} + k^{l-1/l} + k^{(2l-3)/2l} + ... + k^{1/l} + k^{1/2l}$.

Доказательство аналогично утверждению 6.

Утверждение 8. Пусть F_q, $q = p^2$ – расширенное конечное поле, M – сообщение и $M = M_1 \| M_2 \| ... \| M_t$, $l - Ch_t(M)$ – l-каскадное универсальное хеширование по кривой Ферма $x^{(q-1)/3} + y^{(q-1)/3} + 1 = 0$ с большим числом точек. Тогда $l - Ch_t(M)$ определяет универсальное семейство хеш функций $\varepsilon - U\left(q^{2l}, q^k, q\right)$, $\varepsilon = 3\sqrt{k^{1/l}}/(\sqrt{2}q)$, $0 < k \le q^l$, со сложностью вычислений $k + k^{(2l-1)/2l} + k^{l-1/l} + k^{(2l-3)/2l} + ... + k^{1/l} + k^{1/2l}$.

Доказательство аналогично утверждению 6.

Параметры многокаскадного универсального хеширования по алгебраическим кривым представлены в таблице.

Схемы каскадного хеширования	Параметры универсального хеширования	Оценки сложности вычислений
$Ch_t(M)$, $PSh_q - PSh_q$	$\varepsilon - U\left(q^2, q^k, q\right)$, $\varepsilon = k^{1/2}/q$, $0 < k \le q^2$	$k + k^{1/2}$
$Ch_t(M)$, $PSh_q - Hh_q$	$\varepsilon - U\left(q^2\sqrt{q}, q^k, q\right)$, $\varepsilon = (2k)^{1/3}/q$, $0 < k \le q\sqrt{q}/2$	$k + k^{2/3}/2^{1/3} + k^{1/3}/2^{1/6}$
$Ch_t(M)$, $PSh_q - Fh_q$	$\varepsilon - U\left(q^3, q^k, q\right)$, $\varepsilon = (9k/2)^{1/3}/q$, $0 < k \le q\sqrt{q}$	$k + k^{2/3}/2^{1/3} + k^{1/3}/2^{1/6}$
$l - Ch_t(M)$, $PSh_q - PSh_q - ...$	$\varepsilon - U\left(q^l, q^k, q\right)$, $\varepsilon = k^{1/l}/q$, $0 < k \le q^l$	$k + k^{l-1/l} + k^{l-2/l} + ... + k^{1/l}$
$l - Ch_t(M)$, $Hh_q - Hh_q - ...$	$\varepsilon - U\left(q^{l+l/2}, q^k, q\right)$, $\varepsilon = \sqrt{2k^{1/l}}/q$, $0 < k \le q^l$	$k + k^{(2l-1)/2l} + k^{l-1/l} + k^{(2l-3)/2l} + ...$ $+ k^{1/l} + k^{1/2l}$
$l - Ch_t(M)$, $Fh_q - Fh_q - ...$	$\varepsilon - U\left(q^{2l}, q^k, q\right)$, $\varepsilon = 3\sqrt{k^{1/l}}/(\sqrt{2}q)$, $0 < k \le q^l$	$k + k^{(2l-1)/2l} + k^{l-1/l} + k^{(2l-3)/2l} + ...$ $+ k^{1/l} + k^{1/2l}$

Выводы

1. Каскадное хеширование эффективно увеличивает размер хешируемых данных и выравнивает вероятность коллизии с изменением длины данных.

2. Применение многокаскадного универсального хеширование $l - Ch_t(M)$ с одной и той же функцией хеширования на всех каскадах в $\sqrt[l]{k}$ раз уменьшает вероятность коллизии.

Наименьшая вероятность коллизии реализуется в схеме $Hh_q - Hh_q$. Вычисления в поле ~64 бит обеспечивает доказуемую стойкость $P_{кол} < 2^{-57}$ для данных длиной до нескольких Гбт. Затраты по ключам в схеме $Hh_q - Hh_q$ являются наименьшими 124 бит.

3. Каскадное хеширование $PSh_q - PSh_q$ по вероятности коллизии совпадает с однокаскадным хешированием по максимальным плоским кривым и кривой Ферма или Гурвица с большим числом точек и несколько проигрывает $Hh_q - Hh_q$ хешированию. По числу вычислений $PSh_q - PSh_q$ хеширование совпадает с хешированием Hh_q и Fh_q. Преимуществом по сравнению с хешированием по кривой Эрмита являются вычисления в простом поле, в отличие от вычислений в квадратичном поле для Hh_q. Затраты по ключам совпадают с затратами на Fh_q хешированием и определяются значением квадрата поля вычисления.

4. Двухкаскадные схемы хеширования $PSh_q - Hh_q$, $PSh_q - Fh_q$, по вероятности, коллизии являются эквивалентными. Чуть проигрывает хеширование $PSh_q - Fh_q$ и с увеличением размерности поля отличие от $PSh_q - Hh_q$ хеширования становится меньше. Хеширование $PSh_q - Hh_q$ имеет наименьшие затраты по ключам и наименьшую сложность вычислений, для первого каскада в простом поле по одной рациональной функции и по второму каскаду в квадратичном поле по двум рациональным функциям. Затраты по ключам в корень квадратный от размерности поля вычислений больше по сравнению с однокаскадным Fh_q хешированием по кривой Ферма.

Список литературы: 1. *Wegman M. N.* New hash functions and their use in authentication and set equality // Wegman M. N., Carter J. L. // J. Computer and System Science. – 1981. – V. 22. – P. 265-279. 2. *Stinson D.* Universal hashing and authentication codes / Stinson D. // Design, Codes and Cryptography. – 1994. – V. 4. – P.369-380. 3. *Rogaway P.* Bucket hashing and its application to fast message authentication / *Rogaway P.* // rogaway.cs.ucdavis.edu. 4. *Халимов Г.З.* Каскадное универсальное хеширование с использованием АГК кодов / *Халимов Г.З., Иохов А.Ю.* // Восточно-европейский журнал передовых технологий. – Х., 2005. – Вып. 2/2(14). – С. 111–119. 5. *Халимов Г.З.* Багатократне універсальне хешування // Халимов Г.З. // Спеціальні телекомунікаційні системи та захист інформації : Зб. наук. праць. - Київ : ДССЗ та ЗІ. – 2010. – Вип. 2(18). – С.43-49. 6. *Халимов Г.З.* Универсальное хеширование по кривой Сузуки / Г.З. Халимов, Е.В. Котух // Прикладная радиоэлектроника. – Харьков : ХНУРЭ. – 2011. – Т.10, № 2. – С.80-86.

Харьковский национальный университет радиоэлектроники

Поступила в редколлегию 15.08.2011