УДК 681.3.06: 519.248.681

КОЛЛИЗИОННЫЕ ОЦЕНКИ УНИВЕРСАЛЬНОГО ХЕШИРОВАНИЯ НА ОСНОВЕ СХЕМ С АЛГЕБРАИЧЕСКИМИ КОДАМИ

Г.З. ХАЛИМОВ

Рассматриваются схемы универсальных хеш функций на основе традиционных алгебраических и алгеброгеометрических кодов. Получены асимптотические оценки вероятности коллизий для схем универсального хеширования.

The paper considers some schemes of universal hash functions on the basis of traditional algebraic and algebraic and geometrical codes. Asymptotic estimates of a collision probability for universal hashing schemes are obtained.

Возможность применения алгеброгеометрических кодов для целей универсального хеширования впервые была предложена группой авторов Bierbrauer J., Johansson T., Kabatianskii G., Smeets B. [1] и активно исследовалась, прежде всего, в работах Bierbrauer J.

Проблематика построения универсального семейства хеш функций по Картеру-Вегману на основе схем помехоустойчивого кодирования в представлении Bierbrauer J., Johansson T., Kabatianskii G., Smeets B. заключается в выборе алгебраического кода с требуемыми параметрами. Параметры скорость кода и относительное кодовое расстояние входят в оценку вероятности коллизии для схемы универсального хеширования. Оценки вероятности коллизии для многих практических схем хеширования на основе кодов Рида-Соломона, Эрмита, Сузуки, каскадного кодирования в силу комбинаторных свойств алгебраических схем хеширования известны [2–3]. Актуальным является получение асимптотических оценок для наилучших алгебраических и алгеброгеометрических кодовых схем хеширования, что позволит выявить их ограничения. С этой целью в разделе 1 приводятся определение и свойства схем универсального хеширования с использованием алгебраического кодирования. В разделе 2 представлены результаты по оценке асимптотических параметров универсальных схем хеширования с использованием алгеброгеометрических кодов.

1. ОПРЕДЕЛЕНИЕ И СВОЙСТВА СХЕМ УНИВЕРСАЛЬНОГО ХЕШИРОВАНИЯ С ИСПОЛЬЗОВАНИЕМ АЛГЕБРАИЧЕСКОГО КОДИРОВАНИЯ

Асимптотические характеристики вероятности коллизии универсального хеш семейства

Связь между универсальным семейством хеш-функций и кодовыми схемами устанавливает следующая теорема [1].

Теорема 1. Если существует $(n,k,d)_q$ код, тогда существует $\varepsilon(1-\frac{d}{n}) - U(n;q^k,q)$ универсальное семейство хеш- функций, где

$$\varepsilon = (1-\frac{d}{n}). \qquad (1)$$

Справедливо и обратное утверждение. Если $\varepsilon - U(N;\ n,m)$ – хеш семейство, тогда существует $(N,n,N(1-\varepsilon))$ код.

Вычисление хеш значений с использованием $[n,k,d]_q$ линейного кода заключается в отображении q-ичного кодового слова, соответствующего передаваемому сообщению, в значение кодового символа, порядковый номер которого определяется ключевыми данными. Хеш функцию представим следующим определением.

Определение 1. Пусть G – порождающая матрица линейного $[n,k,d]_q$ кода

$$G = \begin{vmatrix} f_0 \\ f_1 \\ \vdots \\ f_{k-1} \end{vmatrix} = \begin{vmatrix} f_0(P_1) & f_0(P_2) & \dots & f_0(P_{n1}) \\ f_1(P_1) & f_1(P_2) & \dots & f_1(P_n) \\ \vdots & \vdots & \dots & \vdots \\ f_{k-1}(P_1) & f_{k-1}(P_2) & \dots & f_{k-1}(P_n) \end{vmatrix}. \qquad (2)$$

Хеш значение для сообщения $m=(m_0,m_1,\dots, m_{k-1})$, $m_i \in F_q$ в точке $P_j \in F_q$ определяется выражением

$$h_{P_j}(m) = \sum_{i=0}^{k-1} f_i(P_j) m_i\ ,\ h_{P_j}(m) \in F_q\ . \qquad (3)$$

Коллизионные характеристики определяются теоремой 1.

Рассмотрим свойства универсального хеширования с кодовой схемой.

Утверждение 1. Пусть $(1-\frac{d}{n}) - U(n;\ q^k,q)$ универсальное семейство хеш-функций, образованное $(n,k,d)_q$ линейным кодом. Тогда хеш семейство является также линейным

$$h(x_1+x_2) = h(x_1)+h(x_2).$$

Этот результат следует прямо из определения хеш функций с использованием кодовых схем.

Утверждение 2. Пусть $\varepsilon - U(n;q^k,q)$ универсальное семейство хеш-функций, образованное алгебраическим $(n,k,d)_q$ кодом, где $\varepsilon = (1-\frac{d}{n})$.

а) Если $(n,k,d)_q$ есть линейный код с единицей, то есть содержит кодовые слова константы $c(x) = a1(x)$, $\alpha \in F_q$, тогда справедливы отношения для вероятности успеха имитационной атаки

$$\frac{1}{q} \le P_{\text{им}} \le 1 \qquad (4)$$

и вероятности подмены

$$P_{\text{под}} \le 1. \quad (5)$$

b) Если $(n,k,d)_q$ – линейный код без единицы, тогда

$$\frac{1}{q} \le P_{\text{им}} \le \varepsilon, \quad (6)$$

$$P_{\text{под}} \le \varepsilon. \quad (7)$$

Верхняя граница вероятности имитации MAC кода по ключу определяется максимальным значением $P_{\text{им.Кл}}$ по всему пространству сообщений и значение вероятности ограничивается соотношением (8) вида

$$P_{\text{им.Кл}} \le \max \frac{\left|\{f \in H : y = f(x)\}\right|}{|H|}, (x,y) \in A \times B, \quad (8)$$

где числитель определяется мощностью множества хеш-функций, которые для сообщения x производят одинаковое MAC значение. Если линейный код содержит единицу, всегда найдётся сообщение x, для которого кодовое слово есть константа $c(x) = a$ и, следовательно, $\left|\{f \in H : y = f(x)\}\right| = |H|$, и справедливо $P_{\text{им.Кл}} \le 1$.

Рассмотрим $(n,k,d)_q$ код, который не содержит слова, все символы которых равны константе. Такой код можно получить из линейного путем удаления из порождающей матрицы единичной строки. Тогда наибольшее число одинаковых элементов в кодовом слове $y \in F_q$ не будет превышать значения $n-d$ и, следовательно, $\left|\{f \in H : y = f(x)\}\right| = n-d$, и справедлива следующая оценка сверху

$$P_{\text{им.Кл}} \le \frac{n-d}{n} = \varepsilon. \quad (9)$$

Вероятность имитации по MAC значению $P_{\text{им.MAC}}$ определяется выражением (10)

$$P_{\text{им.MAC}} = \frac{1}{\left|\{y \in B : y = f(x)\}\right|}, (x,y) \in A \times B, f \in H. \quad (10)$$

Если не известно распределение MAC значений по пространству сообщений, тогда справедлива нижняя граница

$$P_{\text{им.MAC}} \ge \frac{1}{|B|} = q^{-1}. \quad (11)$$

Для линейного кода $(n,k,d)_q$ без единицы верхняя оценка числа различных символов в кодовом слове определяется соотношением

$$\mu \le \frac{n}{n-d}. \quad (12)$$

Отсюда следует выражение для вероятности имитации по MAC значению

$$P_{\text{им.MAC}} \le \frac{1}{\left|\{y \in B : y = f(x)\}\right|} = \frac{n-d}{n} = \varepsilon. \quad (13)$$

Выражение для вероятности подмены с использованием формулы полной вероятности и статистики благоприятных исходов имеет вид

$$P_{\text{под}} = \frac{\left|\{f \in H : y = f(x), y' = f(x')\}\right|}{\left|\{f \in H : y = f(x)\}\right|},$$

$$(x,y), (x',y') \in A \times B, x \ne x', f \in H. \quad (14)$$

Верхняя граница вероятности навязывания путем подмены сообщений и аутентификаторов определяется максимальной вероятностью успеха для всех пар сообщений при равновероятном выборе ключа.

Для линейного кода с единицей существует кодовое слово с одним и тем же значением символов. В этом случае верхняя граница для вероятности подмены в силу соотношения (14) будет равна единице $P_{\text{под}} \le 1$.

Для линейного кода без единицы наибольшее число совпадений кодовых символов определяется значением $n-d$ и вероятность навязывания путем подмены сообщений ограничивается значением ε

$$P_{\text{под1}} \le P_{\text{кол}} = \max_{x,x' \in A} \frac{\left|\{f \in H : y = f(x), y = f(x')\}\right|}{\left|\{f \in H : y = f(x)\}\right|} = \varepsilon,$$

$$y \in B, x \ne x', f \in H. \quad (15)$$

При условии подмены второго рода, когда $y \ne y'$ вероятность подмены также не превышает ε. В силу линейности кода, разность кодовых слов является также кодовым словом и число одинаковых кодовых символов не может превышать $n-d$. Отсюда следует

$$P_{\text{под2}} \le \max_{x,x' \in A} \frac{\left|\{f \in H : y = f(x), y' = f(x')\}\right|}{\left|\{f \in H : y = f(x)\}\right|} = \frac{n-d}{n} = \varepsilon,$$

$$y, y' \in B, x \ne x', y \ne y' f \in H. \quad (16)$$

что завершает доказательство.

Асимптотические характеристики вероятности коллизии $\varepsilon - U(n;q^k,q)$ универсального хеш семейства определяются следующими утверждениями.

Утверждение 3. Асимптотическая граница вероятности коллизии для хеш класса с использованием алгебраического $(n,k,d)_q$ кода при $n,k,d \to \infty$, $k/n \to R$ и $d/n \to \delta$ имеет вид

$$\frac{1}{q} \le P_{\text{кол}} \le \varepsilon \le \frac{1}{q} + \frac{\sqrt{2R(q-1)\ln q}}{q}. \quad (17)$$

Значение ε для хеширования на основе кодовых схем определяется отношением $1-\frac{d}{n}$. Нижняя оценка для ε будет определяться верхней границей для кодового расстояния. Воспользуемся границей Плоткина. Для любого $(n,k,d)_q$ кода справедливо

$$\frac{d}{n} \le \frac{q^k(q-1)}{(q^k-1)q}. \qquad (18)$$

Отсюда для системы $(n_i,k_i,d_i)_q$ кодов, при $k_i \to \infty$ получим

$$\varepsilon \ge 1 - \frac{(q-1)}{q} = \frac{1}{q}. \qquad (19)$$

Для вывода верхней оценки ε воспользуемся нижней границей Варшамова-Гильберта, которая имеет вид

$$R(\delta) = 1 - H(\delta) =$$
$$= 1 - \delta\log_q(q-1) + \delta\log_q\delta + (1-\delta)\log_q(1-\delta).$$

Для оценки поведения ε при больших

$$\delta \to \frac{(q-1)}{q}$$

используем известную асимптотику

$$R\left(\frac{q-1}{q} - x\right) = \frac{q^2}{2(q-1)\ln q}x^2 + o\left(x^2\right),$$

где $x = (q-1)/q - \delta$.

Отбросив малые второго порядка $o\left(x^2\right)$, получим

$$R = \frac{q^2}{2(q-1)\ln q}x^2$$

и $$\delta = \frac{q-1}{q} - \frac{\sqrt{2R(q-1)\ln q}}{q},$$

отсюда $$\varepsilon \le \frac{1}{q} + \frac{\sqrt{2R(q-1)\ln q}}{q}.$$

Это завершает доказательство.

Утверждение 4. Асимптотическая граница вероятности коллизии для хеш-класса с использованием алгебраического $(n,k,d)_q$ кода при $n \to \infty$ и фиксированных k,q имеет вид

$$1 - \frac{q^k(q-1)}{(q^k-1)q} \le P_{\text{кол}} \le \varepsilon \le \frac{k-1}{q}. \qquad (20)$$

Как и ранее нижняя оценка для ε определяется верхней границей Плоткина для кодового расстояния

$$\varepsilon = 1 - \frac{d}{n} = 1 - \frac{q^k(q-1)}{(q^k-1)q}. \qquad (21)$$

Граница существования универсальных хеш-классов с нижней оценкой для ε получается из кодов Рида-Маллера первого порядка с параметрами

$$\left[\frac{q^m-1}{(q-1)}, m, q^{(m-1)}\right]_q. \qquad (22)$$

Для k=1 получим $\varepsilon = 0$, что соответствует тривиальному случаю $(q,1q)_q$ кода с вычислением хеш значений $h_x(m)=mx$ в конечном поле $x,m \in F_q$.

Случай k=2 соответствует кодам Рида-Маллера с параметрами $[(q+1),2,q]_q$ и $\varepsilon = \frac{1}{q+1}$.

При малых значениях $k>2$ и достаточно больших q имеем

$$\varepsilon \ge 1 - \frac{q^k(q-1)}{(q^k-1)q} =$$
$$= \frac{1}{1 + q\left[\frac{1}{1+q^{-1}+q^{-2}+\ldots+q^{-(k-2)}}\right]} \approx$$
$$\approx \frac{1}{1+q\left(\frac{1}{1+q^{-1}}\right)} \approx \frac{1}{q},$$

что сводится к нижней оценке предыдущей асимптотической границы.

Для вывода верхней границы воспользуемся характеристиками кодов БЧХ. Известно, что q-ичный примитивный код БЧХ имеет параметры

$$\left[n = q^m - 1, k \ge n - m\left(d - 1 - \left\lfloor\frac{d-1}{q}\right\rfloor\right), \ge d\right]_q,$$

$$d = 2,3,\ldots,1 + \frac{nq}{m(q-1)}.$$

Из оценки для размерности кода получим

$$k \ge n - m(d-1);$$

$$\frac{n-k+m}{m} \le d;$$

$$\varepsilon = 1 - \frac{d}{n} \le 1 - \frac{n-k+m}{mn} = 1 - \frac{1}{m} + \frac{k}{mn} - \frac{1}{n}.$$

Составляющая $1 - \frac{1}{m}$ в выражении для ε принимает наименьшее значение при $m=1$. Это соответствует случаю РС кодов. Для расширенного РС кода с параметрами $[n=q,k,n-k+1]q$, получим итоговое отношение $\varepsilon \le \frac{k-1}{q}$.

Выводы. Для снижения вероятности коллизии в схемах универсального хеширования на основе алгебраических кодов необходимо использовать низкоскоростные длинные q–ичные коды в поле большой размерности $q>>1$ и с большим кодовым расстоянием $1 - \frac{d}{n} << 1$.

Практический интерес представляют алгебро-геометрические коды.

2. ОЦЕНКА АСИМПТОТИЧЕСКИХ ПАРАМЕТРОВ УНИВЕРСАЛЬНЫХ СХЕМ ХЕШИРОВАНИЯ С ИСПОЛЬЗОВАНИЕМ АЛГЕБРОГЕОМЕТРИЧЕСКИХ КОДОВ

Впервые алгеброгеометрический подход к построению кодов по алгебраическим кривым был предложен Гоппой В.Д. в 1981 году [6]. В рабо-

тах [4,5] показано, что по алгебраическим кривым можно строить коды с очень хорошими асимптотическими свойствами. Доказано существование бесконечных серий q-ичных линейных кодов, параметры которых (при $q=2^{2n}>49$ и $N\to\infty$) лежат выше границы Варшамова – Гильберта.

Результаты по алгеброгеометрическим кодам обобщаются в следующей теореме [7].

Теорема 2. Пусть, зафиксированы: гладкая алгебраическая кривая X рода g над F_q, наборы точек

$$P=\{P_1,P_2,...,P_n\}\subseteq X(F_q),$$

$$Q=\{Q_1,Q_2,...,Q_s\}\subseteq X(F_q),\ Q\cap P=\varnothing$$

(условие не существенно и снимается использованием пучковой конструкции), эффективный дивизор $D=\sum u_Q Q$ степени $\deg(D)=\sum u_Q$ и $\deg(D)>2g-2$. С дивизором D ассоциируется линейное пространство $L(D)$ из рациональных функций $f_0,f_1,...,f_{k-1}$ на X как множество всех функций f таких, что порядок f в каждой точке Q удовлетворяет условию $div(f)\ge -u_Q$.

Тогда линейный код C над F_q определяется как

$$C=\{f(P_1),...,f(P_n)\,|\,f\in L(D)\}.$$

Причем алгеброгеометрический код имеет параметры

$$[n,\deg(D)-g+1,d],\ d\ge n-\deg(D),$$

а двойственный к нему код также является алгеброгеометрическим с параметрами

$$[n,n-\deg(D)+g-1,d^{\perp}],\ d^{\perp}\ge n-\deg(D)-2g+2.$$

Алгеброгеометрическая кодовая конструкция для универсального хеширования определена в [2] и свойства представлены следующей теоремой.

Теорема 3 [2]**.** Пусть задана алгеброгеометрическая кривая X над F_q с $N+1$ рациональными точками. Пусть P является одной из этих рациональных точек с порядками полюса u_k, где $u_0=0<u_1<u_2...$ и $u_k<N$. Тогда алгеброгеометрический код размерности k и минимальным расстоянием $d\ge N-u_k$ образует универсальный класс хеш-функций $\varepsilon-U(N,q^k,q)$, где $\varepsilon\le\frac{u_k}{N}$.

Алгебраические кривые являются проективными многообразиями размерности $\dim X(F_q)=2$. Существуют тонкие алгеброгеометрические конструкции (модулярные кривые, башни Гарсии-Штихтенота), которые определяются как проективные многообразия размерности $\dim X(F_q)\ge 3$, с большим родом и числом точек. Граница Дринфельда-Влэдуца определяет оценку для отношения максимального числа точек $N_g(q)$ на кривой к её роду g

$$\limsup_{g\to\infty}\frac{N_g(q)}{g}\le\sqrt{q}-1. \qquad (23)$$

Следующее утверждение определяет основную алгеброгеометрическую границу для вероятности коллизии $\varepsilon-U(N,q^k,q)$ универсального хеш семейства.

Утверждение 5. Основная алгеброгеометрическая граница для вероятности коллизии $\varepsilon-U(N,q^k,q)$ хеш-класса, построенного с использованием алгеброгеометрических кодовых конструкций, имеет вид

$$P_{\text{кол}}\le\ \varepsilon=\frac{1}{\sqrt{q}-1}+R, \qquad (24)$$

где R – скорость кода.

Для алгеброгеометрических кодов основная асимптотическая граница Цфацмана-Влэдуца-Цинка при $N\to\infty$ имеет вид

$$R=1-\frac{1}{\sqrt{q}-1}-\delta, \qquad (25)$$

где δ – относительное кодовое расстояние, R – скорость кода.

Применение оценки (25) приводит к основной алгеброгеометрической границе вероятности коллизии

$$\varepsilon=1-\frac{d}{N}=1-\delta=\frac{1}{\sqrt{q}-1}+R.$$

Основная алгеброгеометрическая граница вероятности коллизии лежит ниже асимптотической границы, построенной с использованием кодовой границы Варшамова-Гильберта для практически значимых значений $q\ge 49$.

Рассмотрим алгебраические многообразиями размерности $\dim X(F_q)=2$. Справедливо следующее утверждение.

Утверждение 6. Асимптотическая граница вероятности коллизии для $\varepsilon-U(N,q^k,q)$ хеш-класса, построенного на основе алгеброгеометрического $(N,k,d)_q$ кода по максимальным кривым над большим алфавитом и фиксированных $k\le 2g$ и q, имеет вид

$$1-\frac{q^k(q-1)}{(q^k-1)q}\le P_{\text{кол}}\le\varepsilon<\frac{\sqrt{2k}}{q}. \qquad (26)$$

Нижняя оценка ε определяется известной верхней границей Плоткина для кодового расстояния алгебраических кодов

$$\varepsilon=1-\frac{d}{n}=1-\frac{q^k(q-1)}{(q^k-1)q}.$$

Для вывода верхней границы используем оценку для кодового расстояния алгеброгеометрического кода размерности k

$$d\ge N-u_k,$$

где u_k – максимальный порядок полюсов рациональных функций $f_0,f_1,...,f_{k-1}$ на алгебраической кривой $X(F_q)$, с точками $P=\{P_1,P_2,...,P_n\}\subseteq X(F_q)$, $Q=\{Q_1,Q_2,...,Q_s\}\subseteq X(F_q)$, $Q\cap P=\varnothing$, образующих

базис линейного пространства $L(D)$, который ассоциированный с дивизором D.

Верхняя оценка вероятности коллизии ε имеет вид

$$\varepsilon = 1 - \frac{d}{N} \le \frac{u_k}{N}. \qquad (27)$$

Порядок базисных функций f в каждой точке Q удовлетворяет условию $div(f) \ge -u_k$. Рассмотрим максимальные кривые, число точек которых лежит на границе Серра

$$N = q + 1 + 2g\sqrt{q}, \qquad (28)$$

где g – род кривой и известно, что для максимальной кривой $g \le \frac{q - \sqrt{q}}{2}$. Пусть степень алгебраической кривой равна *s*. Род максимальной кривой определяется выражением

$$g = \frac{(s-1)(s-2)}{2}.$$

Отсюда получим, что

$$s = \frac{\sqrt{8g+1}}{2} + \frac{3}{2}. \qquad (29)$$

Порядки полюсов u_i, $u_0 = 0 < u_1 < u_2 \ldots$, рациональных функций $f_0, f_1, \ldots, f_{k-1}$ на алгебраической кривой $X(F_q)$ в простейшем случае образуют аддитивную группу вида

$$u_i = lu_1 + mu_2, \qquad (30)$$

где $l = 0,1,2,\ldots$, $m = 0,1,2,\ldots$, $u_1 = s - 1$, $u_2 = s$.

Для $k \le 2g$, значение полюса рациональной функции f_k, в силу условия $u_0 = 0 < u_1 < u_2 \ldots$ и соотношения (30) имеет верхнюю оценку

$$u_k \approx s\sqrt{2k}. \qquad (31)$$

Подставляя соотношения (29), (31) в (27), получим

$$\varepsilon \le \frac{u_k}{N} = \frac{\sqrt{2k}\left(\sqrt{8g+1}+3\right)}{2N}. \qquad (32)$$

Для алгебраической кривой максимального рода имеем $g = \frac{q - \sqrt{q}}{2}$ и $N = q\sqrt{q} + 1$. При больших q, с учетом отношения (32), получим искомую асимптотику для ε

$$\varepsilon \le \frac{u_k}{N} \le \frac{\sqrt{2k}\left(\sqrt{8\left(q - \sqrt{q}/2\right)+1}+3\right)}{2\left(q\sqrt{q}+1\right)} \approx$$
$$\approx \frac{\sqrt{2k}\sqrt{q}}{q\sqrt{q}} + \frac{3\sqrt{2k}}{2q\sqrt{q}} \approx \frac{\sqrt{2k}}{q}.$$

Выводы. Из результата утверждения 6 следует, что алгеброгеометрические коды над большим алфавитом являются более предпочтительными по сравнению с РС кодами, так как обеспечивают в универсальных схемах хеширования при фиксированных k и q меньшее значение вероятности коллизии.

Литература.

[1] *Bierbrauer J., Johansson T.,Kabatianskii G., Smeets B.* On families of hash functions via geometric codes and concatenation.//Advances in Cryptology-CRYPTO '93 Proceedings, Springer-Verlag, 331-342 (1994).

[2] *Jurgen Bierbrauer.* Authentication via algebraic-geometric codes. URL http://www.math.mtu.edu/~jbierbra/ potpap.ps.

[3] *Халимов Г.З., Кузнецов А.А.* Аутентификация с применением алгеброгеометрических кодов. //Радиотехника. Всеукр. межвед. науч.-техн. сб. 2001. Вып. 120. С. 103-109.

[4] *Халимов Г.З., Иохов А.Ю.* Каскадное универсальное хеширование с использованием АГК кодов // Восточно-европейский журнал передовых технологий. – Х., -2005. – Вып. 2/2 (14). –С. 155-119.

[5] *Халимов Г.З., Иохов А.Ю.* Аутентификация с применением эрмитовых кодов// Вестник ХПИ. – Х., – 2005. НТУ «ХПИ». – Вып. 9. – С. 26-32.

[6] *Гоппа В.Д.* Коды на алгебраических кривых //Докл. АН СССР. – 1981. – Т.259. № 6. – С. 1289-1290.

[7] *Цфасман М.А.* Коды Гоппы, лежащие выше границы Варшамова - Гилберта. //Проблемы передачи информации. – 1982 – Т.18. №3 – С 3-6.

Поступила в редколлегию 23.09.2009

Халимов Геннадий Зайдулович, кандидат технических наук, доцент кафедры БИТ ХНУРЭ. Область научных интересов: методы и средства высокоскоростной аутентификации данных.