УДК 621.322

А.А. КУЗНЕЦОВ[1], В.Е. ЧЕВАРДИН[2]

[1]*Харьковский университет Воздушных Сил, Украина*
[2]*Полтавский военный институт связи, Украина*

МОДЕЛЬ ОЦЕНКИ ПАРАМЕТРОВ УНИВЕРСАЛЬНОЙ ХЕШИРУЮЩЕЙ ФУНКЦИИ

Представлена математическая модель оценки параметров универсальной хеш-функции, которая позволяет оценить вероятность коллизий при заданном объеме хешируемых данных. Рассмотрены различные случаи соотношения параметров полученной модели. Представлены рекомендации по ее использованию.

хеш-функция, оценка вероятности коллизий, объем хешируемых данных, универсальный класс

Введение

Постановка задачи. Оценка параметров криптографических систем является одной из главных исследовательских задач в области криптографии. Это объясняется необходимостью качественной характеристики криптосистем и представления требуемых гарантий безопасности информации их владельцу.

Особое место среди современных криптографических систем занимают хеш-функции, область применения которых очень широка (схемы ЦП, MAC-коды, схемы аутентификации и т.д.). Известно [1], что главным показателем оценки эффективности схем хеширования является вероятность коллизии, ее точная оценка влияет на правильный выбор механизмов обеспечения безопасности информации и позволяет прогнозировать ее риски в различных случаях. Для определенного класса хеш-функций (универсальных) этот показатель теоретически доказуем, что позволяет обеспечить бóльшую уверенность в стойкости схемы хеширования к коллизиям.

Оценка эффективности универсальных классов хеш-функций, доказательства крайних (верхних) границ параметров (граница Гилберта, Мак-уильямса и Слоуна) имеет важное научное значение и практическое применение для криптографии. В связи с этим **актуальным** является разработка моделей оценки параметров универсальных классов хеш-функций.

Существующие границы схем аутентификации данных (схем хеширования). Известной атакой на схемы хеширования [2] является атака встречи по средине. Сложность этой атаки оценивается как $2^{(l+1)/2}$ операций вычисления хеш-значений, а ее эффективность зависит от интенсивности обмена данными, т.е. от объема пар хешируемый текст – открытый текст, которыми может обладать криптоаналитик. Гилбертом, Мак-уильямсом и Слоуном [3] была получена нижняя граница вероятности успешного обмана

$$P_d \geq 1/\sqrt{|\varepsilon|},$$

где $|\varepsilon|$ – общее число правил кодирования (число ключей хеширования). При наличии этих границ для оценки схем аутентификации возникают вопросы: каков может быть максимальный объем хешируемых данных при фиксированной вероятности коллизий и какова может быть минимальная вероятность коллизий при фиксированном объеме хешируемых данных? Однако математических выражений, позволяющих связать объем хешируемых данных с вероятностью коллизий в доступных источниках не найдено.

В связи с этим **целью данной работы** является разработка математической модели универсальной схемы хеширования, которая позволит: во-первых, связать объем хешируемых данных с вероятностью коллизий, а во-вторых, получить новых подход к

оценке числа существующих классов универсальных хеш-функций.

Распределение значений хеш-функции с позиции матричного и векторного представления

Впервые универсальные ключевые хеш-функции были предложены в работах Картера и Вегмана [3]. Рассмотрим существующие подходы к построению универсальных хеш-функций с использованием определений, введенных в этих работах.

Согласно определений [4, 5] хеш-функцию представляют в виде универсального класса функций $\varepsilon - U(N, n, r)$, где N – количество функций отображения множества открытых текстов $\sum^n$ мощности n в множество хеш-кодов $\sum^r$ мощности r. Причем для двух различных элементов $M_1, M_2 \in \sum^n$ существует не больше, чем $N \cdot \varepsilon$ функций $f \in H$ таких, что $f(M_1) = f(M_2)$. Примерами являются классы хеш-функций, построенные на основе полиномиальных функций, на основе РС-кодов, на основе ортогональных массивов, детальное исследование которых представлено в [5, 6].

Очевидным фактом является представление значений хеш-функции в виде матрицы (совокупности векторов).

Представим распределение значений хеш-функции относительно ключей и хешируемых текстов в виде матрицы. Пусть H – матрица, состоящая из q строк и q столбцов, элементы которой принимают одно из g значений, равное q (табл. 1). Каждая функция, параметризованная ключом ki, связана со строкой и определяет правило отображения элементов Mq (номеров столбцов матрицы) в hq (собственные значения матрицы).

Таблица 1

Матрица H

$k\backslash M_i$	M_1	M_2	…	M_q
k_1	h_1	h_2	…	h_r
k_2	h_2	h_1	…	h_{r-1}
…	…	…	…	h_{r-2}
k_q	$h_{1+(q-1)}$	$h_{2+(q-1)}$	…	$h_{r+(q-1)}$

Наилучшим (желаемым) распределением хеш-значений в такой матрице является случай, когда в каждом ее столбце либо строке значений хеш-функции встречается один раз. Такое отображение соответствует классу $\varepsilon - U(N, n, r)$ при $N = n = r = q$, $\varepsilon = 0$, и может быть получено с использованием произвольного автоморфизма поля, что не требует доказательств. Представим распределение хеш-значений для случая $n > r$ в виде матрицы $H1$ размерности $q \times lq$ (табл. 2).

Таблица 2

Матрица $H1$

$k\backslash M_i$	M_1	M_2	…	M_q	M_{q+1}	…	M_{2q}	M_{2q+1}	…	M_{lq}
k_1	h_1	h_2	…	h_r	h_2	…	h_r	h_1	…	h_r
k_2	h_2	h_1	…	h_{r-t}	h_1	…	h_1	h_2	…	H_{r-1}
…	…	…	…	…	…	…	…	…	…	…
k_q	h_r	h_{r-1}	…	h_{r-d}	h_{2+d}	…	h_{r-d}	h_r	…	$h_{r-d)}$

Очевидно, что матрица $H1$ получена путем перестановки элементов в столбцах матрицы H по определенному закону. Нас интересует число функций H_k (ключей), которые дают совпадение значений хеш-функции для фиксированных $M_i \neq M_j$, $H_{k_i}(M_1) = H_{k_j}(M_2)$ при заданном объеме открытых текстов n, т.е. число позиций, в которых совпадают вектора заданного объема. Это позволит определить значение ε (границу вероятности коллизий) для $\varepsilon - U(N, n, r)$ класса хеш-функций при фиксированном n. Для определения этого числа рассмотрим матрицу в виде совокупности векторов базиса $GF(q)$, количество которых соответствует мощности множества открытых текстов.

В матрице H каждый столбец (вектор) $\{h_i\}^j$, $i = 1, q$ отличается друг от друга во всех позициях $\{h_i\}^k \neq \{h_i\}^t$. Определим, сколько существует векторов матрицы $H1$ $q \times lq$, отличающихся в t позициях.

Рассмотрим случай совпадения векторов в одной позиции. Если первый элемент вектора $\{h_i\}^j$ принимает одно из q значений, то второй элемент h_2 может принимать уже из $q-1$ значений, т.е. количество таких векторов равно

$$n = q\,(q-1). \qquad (1)$$

Для совпадения векторов $\{h_i\}^j$ в двух позициях $\{h_i\}^k = \{h_i\}^t$ количество векторов равно

$$n = q(q–1)(q–2).$$

Для случая совпадения в трех позициях n составит

$$n = q(q–1)(q–2)(q–3).$$

Пусть t – количество совпадений элементов любых векторов $\{h_i\}^j$. Индукцией по t получим выражение для определения числа существующих векторов для заданного t:

$$n = q(q–1)(q–2)\,(q–3)\ldots\,(q–t)). \qquad (2)$$

Представим полученное выражение (2) в форме

$$n = \frac{q!}{(q-(t+1))!}. \qquad (3)$$

Для подтверждения справедливости полученного выражения рассмотрим случай $t + 1$:

$$n_t = \frac{q!}{(q-(t+1))!};\ n_{t+1} = \frac{q!}{(q-(t+2))!} = n_t \cdot (q-(k+1)),$$

что соответствует выражению (2) при $t = t + 1$, т.е. сводится к полученному выражению (2).

Для рассмотренного случая число векторов, отличающихся друг от друга во всех позициях, равно

$$n = \frac{q!}{(q-(t+1))!} = \frac{q!}{(q-(0+1))!} = q.$$

Пример. Зафиксируем множество открытых текстов мощности $q = 4$, множество хеш-кодов мощности $r = q = 4$, представляющих матрицу 4×4, подобную матрице H. Из матрицы 4×4 можно получить две матрицы 4×12 (А и В, табл. 3, 4) путем перестановки элементов векторов $\{h_i\}^j$.

Таблица 3

Матрица A

	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	1	2	2	2	3	3	3	4	4	4
2	2	3	4	3	4	1	2	4	1	1	2	3
3	3	4	2	1	3	4	4	1	2	3	1	2
4	4	2	3	4	1	3	1	2	4	2	3	1

Таблица 4

Матрица B

	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	1	2	2	2	3	3	3	4	4	4
2	2	3	4	3	4	1	2	4	1	1	2	3
3	4	2	3	4	1	3	1	2	4	2	3	1
4	3	4	2	1	3	4	4	1	2	3	1	2

Количество полученных векторов в данном случае равно $q! = 4! = 4{\cdot}3{\cdot}2 = 24$. Полученные матрицы отличаются тем, что вектора в каждой из них отличаются друг от друга в $q–1 = 3$ позициях.

Согласно полученного выражения (2), для $t = 1$ количество векторов (одной матрицы), совпадающих в одной позиции, равно

$$n = \frac{q!}{(q-(t+1))!} = \frac{4!}{(4-(1+1))!} = \frac{4!}{2!} = 12,$$

соответствует одной из матриц A, В для t = 2

$$n = \frac{q!}{(q-(t+1))!} = \frac{4!}{(4-(2+1))!} = \frac{4!}{1!} = 24,$$

соответствует матрице, состоящей из матриц А, В.

Таким образом, распределение хеш-значений с параметрами класса хеш-фукций ε–$U(N, n, r)$, $N = r$ можно представить тремя случаями:

1) совокупностью векторов матрицы A либо B;

2) совокупностью векторов двух матриц A и B;

3) совокупность векторов двух матриц A и B с векторами типа {1111}, {1122}.

Выражение (3) соответствует случаю 1. Для произвольного поля $GF(q)$ с учетом этого выражения преобразуем выражение $\varepsilon – U(N, n, r)$, при $N = r = q$, $g = r = q$:

$$\frac{t}{q} - \left(q,\ \frac{q!}{(q-(t+1))!},\right)q. \qquad (4)$$

Выражение (4) позволяет связать значение ε, объем хешируемых данных при фиксированном объеме ключевых данных и хеш-значений. Следовательно, при введенных условиях $N = r$ выражение (4) позволяет определить граничные значения параметров универсальной хеш-функции. Существующие способы ключевого универсального хеширования могут быть лишь максимально приближенными по параметрам к классу (4). Полученные выражения (3, 4) могут использоваться для сравнения и оценки существующих универсальных классов хеширования.

Согласно выражения (4), для $q = 4$ класс универсальных хеш-функций имеет вид

$$\frac{2}{4} - U\left(4,\ \frac{4!}{(4-(2+1))!},\ 4\right) \Rightarrow \frac{1}{2} - U(4,\ 24,\ 4).$$

На рис. 1 представлена зависимость вероятности коллизий от значений q (в двоичном виде) при фиксированном t.

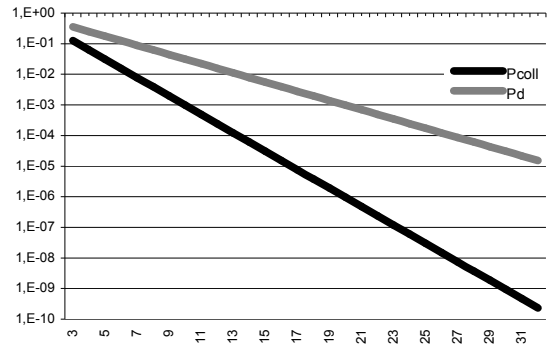

Рис. 1. Зависимость вероятности коллизий от объема хешируемых данных универсальной хеш-функции

На рис. 2 представлена зависимость вероятности коллизий от объема хешируемых данных при фиксированном значении $q = 2^7$ и изменении значений параметра t.

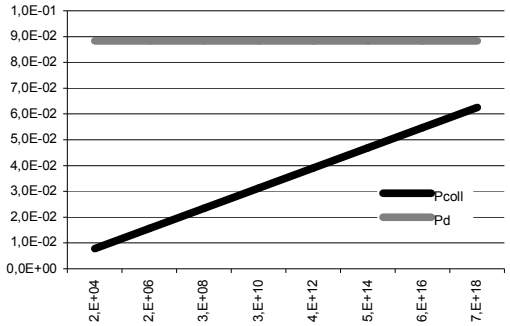

Рис. 2. Зависимость вероятности коллизий от объема хешируемых данных универсальной хеш-функции

Полученные зависимости (рис.1, 2) позволяют получить при фиксированной вероятности коллизий максимально возможный объем хешируемых данных и для заданного значения вероятности коллизий теоретически возможный объем хешируемых данных.

Выводы

Полученная математическая модель универсальной хеш-функции позволяет определить связь значения ε (вероятность коллизий) с объемом хешируемых данных при фиксированном объеме ключей. Это дает возможность теоретически обосновать максимально возможное значение ε для заданного объема хешируемых данных при фиксированном количестве ключей и хеш-кодов. В результате проведенных исследований $\varepsilon - U(N, n, r)$ класса хеш-функций при $N = r$ получена верхняя граница вероятности коллизий $\varepsilon = t/q$ для случая, когда объем хешируемых данных превосходит объем хеш-значений и объем ключевых данных в $\frac{q-1}{(q-(t+1))!}$ раз. Предложенная математическая модель может быть использована для оценки существующих классов универсальных хеш-функций путем сравнения с теоретически возможными значениями вероятности коллизий для заданного объема хешируемых данных либо сравнения с максимально возможным объемом хешируемых данных при фиксированной вероятности коллизий.

Литература

1. Simmons G.I. Authentication theory/coding theory, presented at Crypto'84, Santa Barbara, CA. – 1984. – P. 19-22.

2. Ohta K. and Koyama K. Meet-in-the-Middle Attack on Digital Signature Schemes. In Abstract of AUSCRYPT '90. – 1990. – P. 110-121.

3. Симмонс Г. Дж. Обзор методов аутентификации информации // ТИИЭР. – 1988. – Т. 76, №5. – С. 105-126.

4. Wegman M., Carter L. New hash functions and their use in authentication and set equality // Journal of Computer and System Science. – 1981. – Vol. 22. – P. 265-279.

5. Халимов Г.З., Кузнецов А.А. Аутентификация и универсальное хеширование // Радиотехника. – Х.: ХНУРЭ. –2001. – Вып. 119. – С. 88-94.

6. Халимов Г.З., Кузнецов А.А. Универсальное хеширование на основе кодовых конструкций // Радиотехника. – Х.: ХНУРЭ. – 2001. – Вып. 119. – С. 95-102.

Поступила в редакцию 10.03.2006

Рецензент: д-р техн. наук, проф. Ю.В. Стасев, Харьковский университет Воздушных Сил, Харьков.