УДК 681.3.06

А.А. Кузнецов[1], О.Г. Король[2], В.В. Босько[3], С.П. Евсеев[2]

[1]*Харьковский университет воздушных сил им. И. Кожедуба, Харьков*
[2]*Харьковский национальный экономический университет, Харьков*
[3]*Кировоградский национальный технический университет, Кировоград*

МЕТОДИКА ИССЛЕДОВАНИЯ КОЛЛИЗИОННЫХ СВОЙСТВ КОДОВ АУТЕНТИФИКАЦИИ СООБЩЕНИЙ

Рассматривается схема формирования кодов аутентификации сообщений, основанная на использовании универсального хеширования (UMAC - Message Authentication Code using Universal Hashing). Предлагается методика статистического исследования коллизионных свойств UMAC. В основе предлагаемой методики лежит использование уменьшенной модели кодов аутентификации сообщений (мини-UMAC), построенной посредством масштабирования применяемых преобразований с сохранением их алгебраической структуры.

Ключевые слова: *коды аутентификации сообщений, коллизионные свойства, аутентификация, MAC-код UMAC.*

Постановка проблемы в общем виде и анализ литературы

Эффективным механизмом обеспечения аутентичности и целостности информации являются коды аутентификации сообщений (MAC - коды), в том числе, построенные с использованием универсального хеширования (UMAC – Message Authentication Code using Universal Hashing) [1–7].

Результаты проведенного анализа показали, что UMAC-коды обладают высокими показателями быстродействия и криптографической стойкости [5, 7]. Это достигается применением эффективных схем универсального хеширования и блочного симметричного шифрования. В тоже время, коллизионные свойства UMAC-кодов после применения симметричного шифрования остаются неисследованными, что и определяет актуальность данной работы.

В статье предлагается методика статистического исследования коллизионных свойств UMAC. В основе предлагаемой методики лежит использование уменьшенной модели кодов аутентификации сообщений (мини-UMAC), построенной посредством масштабирования применяемых преобразований с сохранением их алгебраической структуры. Предлагаемая методика позволяет на основе оценки распределения столкновений формируемых образов экспериментально исследовать коллизионные свойства кодов аутентификации сообщений.

Изложение основного материала

Коды аутентификации сообщений UMAC

Схема формирования кодов аутентификации сообщений UMAC использует в своей структуре несколько слоев преобразования, в том числе блочный симметричный шифр (рекомендован к использованию шифр AES) [1–7]. Код подлинности сообщений (обозначим его Tag) по спецификации алгоритма UMAC формируется посредством вычисления следующей функции:

$$Tag = UMAC(K, M, Nonce, Taglen) = Y \oplus Pad,$$

где K – секретный ключ, длина которого $Keylen$ равна стандартной длине секретного ключа используемого блочного симметричного шифра, $Keylen$ принадлежит множеству допустимых значений $\{16, 24, 32\}$ байт); M – информационное сообщение, подлежащее аутентификации, представленное в виде массива-строки, размерностью от одного до 2^{67} бит (2^{64} байт); $Nonce$ – неповторяющееся (для всех вводимых информационных сообщений M) восьмибайтное число; $Taglen$ – целое число из множества допустимых значений $\{4, 8, 12, 16\}$, задающее длину кода подлинности сообщений Tag в байтах; $Y = Hash(K, M, Taglen)$ – функция ключевого универсального хеширования информационного сообщения M с использованием секретного ключа K; $Pad = PDF(K, Nonce, Taglen)$ – функция формирования псевдослучайной подложки (Pad) по введенному значению $Nonce$ и секретному ключу K; «$\oplus$» – побитовое сложение (XOR) результата ключевого хеширования сообщения $Y = Hash(K, M, Taglen)$ и сформированной подложки $Pad = PDF(K, Nonce, Taglen)$, т.е.

$$Tag = Hash(K, M, Taglen) \oplus PDF(K, Nonce, Taglen).$$

Длина хеш-кода Y, подложки Pad и кода Tag принадлежат множеству допустимых значений $\{32, 64, 96, 128\}$ бит. Эти фиксированные значения

$Taglen$ соответствуют случаю формированию кодов подлинности сообщений UMAC – 32, UMAC – 64, UMAC – 96 или UMAC – 128, соответственно.

Таким образом, схема UMAC использует многослойную конструкцию на основе ключевого хеширования $Hash(K, M, Taglen)$ и процедуру формирования псевдослучайной подложки Pad блочным симметричным шифром. Применение универсального хеширования позволяет обеспечить равновероятность формирования хеш-образов для всего множества используемых ключевых данных, формирование псевдослучайной подложки криптографически стойким алгоритмом (например, с использованием блочного симметричного шифра AES) обеспечивает высокую криптостойкость алгоритма UMAC.

В тоже время на сегодняшний день не исследованы коллизионные свойства алгоритма UMAC после применения завершающей процедуры наложения на формируемые хеш-коды $Y = Hash(K, M, Taglen)$ псевдослучайных подложек $Pad = PDF(K, Nonce, Taglen)$.

Методика исследования коллизионных свойств UMAC

В основе предлагаемой методики исследования коллизионных свойств кодов аутентификации сообщений UMAC лежит использование уменьшенных моделей отдельных слоев преобразований и оценка распределения коллизий (столкновений) формируемых образов (кодов).

Применение уменьшенных моделей используемых слоев преобразований позволяет, сохранив алгебраическую структуру криптоалгоритма, проводить исследования основных показателей его эффективности. Этот поход широко используется на сегодняшний день при исследовании криптографических свойств блочных симметричных шифров [8–12]. Коме того, на основе анализа уменьшенных моделей в работах [11, 12] предложен подход к оценке эффективности блочных симметричных шифров в виде вычислительных затрат, требуемых для достижения шифром асимптотических характеристик случайной подстановки.

В настоящей работе предлагается дальнейшее развитие данного направления, состоящее в использовании уменьшенных моделей отдельных слоев преобразований для оценки коллизионных свойств формируемых кодов аутентификации сообщений.

Уменьшенная модель формирования UMAC (мини-UMAC). Схема формирования кодов аутентификации сообщений UMAC использует в своей структуре несколько слоев преобразования, в том числе блочный симметричный шифр (рекомендован к использованию шифр AES). Разрабатываемая уменьшенная модель UMAC должна включать соответствующие слои преобразования с сохранением их алгебраической структуры при выполнении масштабирования до мини-версии. Естественным представляется исследовать коллизионные характеристики формируемых образов (кодов) на каждом из слоев преобразования, в том числе формируемых с помощью блочного симметричного шифра псевдослучайных подложек Pad, проанализировать их влияние на коллизионные свойства кодов аутентификации сообщений уменьшенной модели UMAC.

Выше было показано, что схема формирования кодов UMAC состоит из следующих слоев:

- трехуровневое универсальное хеширование для формирования хеш-кодов $Y = Hash(K, M, Taglen)$;

- криптографическое преобразования с использованием блочного симметричного шифра для формирования псевдослучайной подложки $Pad = PDF(K, Nonce, Taglen)$;

- заключительное преобразование для формирования кодов аутентификации сообщений $Tag = UMAC(K, M, Nonce, Taglen) = Y \oplus Pad$.

Рассмотрим каждый слой схемы формирования кодов аутентификации сообщений UMAC на предмет их масштабирования.

Мини-версию трехуровневого универсального хеширования построим без изменения структуры алгебраических преобразований простым уменьшением размерности блоков обрабатываемых данных в восемь раз.

Соответствующая длина хеш-кода $Y_{\min i}$ уменьшенной модели первого слоя будет кратна 4 битам, его значение сформируем посредством объединения (конкатенации) четырех последовательностей $Y_{\min i L3_i}$

$$Y_{\min i} = Y_{\min i L3_1} \| Y_{\min i L3_2} \| Y_{\min i L3_3} \| Y_{\min i L3_4},$$

где $Y_{\min i L3_i}$ – результат многоуровневого хеширования сообщения уменьшенной модели первого слоя мини-UMAC.

Рассмотрим процесс формирования одного блока $Y_{\min i L3_i}$ (второй уровень хеширования в уменьшенной модели выполнять не будем):

$$Y_{mini L3_i} = Y_{mini L3} = Hash_{mini L3}\left(K_{mini L3_1}, K_{mini L3_2}, Hash_{mini L1}\left(K_{mini L1}, M_{min}\right)\right),$$

где $K_{\min i L1}$, $K_{\min i L3_1}$, $K_{\min i L3_2}$ – ключевые последовательности мини-UMAC,

$Hash_{\min i L1}$ и $Hash_{\min i L3}$ – уменьшенные версии хеширования первого и третьего уровней соответственно.

На первом уровне массив-строка $M_{\min i}$ размерности 32 бита преобразуется функцией $NH(K_{L1}, M_i)$. Эта

строка и является результатом хеширования первого уровня: $Y_{mini\,L1} = NH_{mini}\left(K_{mini\,L1}, M_{mini}\right)$.

Значение функции $NH_{mini}\left(K_{mini\,L1}, M_{mini}\right)$ вычисляется по следующему правилу. Информационный блок $M_{\min i}$ разбивается на восемь четырехбитовых подблоков $M_{mini} = M_{mini_1} \| M_{mini_2} \| ... \| M_{mini_8}$.

Аналогичным образом ключевая последовательность K_{L1} представляется в виде последовательностей из восьми четырехбитовых подблоков: $K_{mini\,L1} = K_{mini\,L1_1} \| K_{mini\,L1_2} \| ... \| K_{mini\,L1_8}$.

После чего (принимая начальное состояние $Hash_{L1} = 0$) выполняются следующие операции:

$$Hash_{mini\,L1} = Hash_{mini\,L1} +_8 ((M_{mini_0} +_4 K_{mini\,L1_0}) \times_8 (M_{mini_4} +_4 K_{mini\,L1_4})),$$

$$Hash_{mini\,L1} = Hash_{mini\,L1} +_8 ((M_{mini_1} +_4 K_{mini\,L1_1}) \times_8 (M_{mini_5} +_4 K_{mini\,L1_5})),$$

$$Hash_{mini\,L1} = Hash_{mini\,L1} +_8 ((M_{mini_2} +_4 K_{mini\,L1_2}) \times_8 (M_{mini_6} +_4 K_{mini\,L1_6})),$$

$$Hash_{mini\,L1} = Hash_{mini\,L1} +_8 ((M_{mini_3} +_4 K_{mini\,L1_3}) \times_8 (M_{mini_7} +_4 K_{mini\,L1_7})),$$

где $+_8$, $+_4$ – операции сложения по модулю 2^8 и 2^4, соответственно; $\times_8$ – операция умножения по модулю 2^8.

В результате вычислений формируется восьмибитное значение $Y_{mini\,L1} = Hash_{mini\,L1}$.

Третий уровень хеширования преобразует поданные на его вход восьмибитные данные $Y_{\min i\,L1}$ в хеш-код $Y_{\min i\,L3}$ длины 4 бита. В качестве ключевых последовательностей выступают $K_{mini\,L3_1}$ и $K_{mini\,L3_2}$ длины 16 и 4 бита соответственно.

Хешируемые данные $Hash_{\min i\,L1}$ и ключевая последовательность $K_{\min i\,L3_1}$ равномерно разбиваются на четыре блока, каждый из которых представляется как целое число $Y_{\min i\,L2_i}$ и $K_{\min i\,L3_{1_i}}$, $i = 1,2,...,4$.

Хеш-значение $Y_{\min i\,L3}$ вычисляется следующим образом:

$$Y_{mini\,L3} = \left(\left(\left(\sum_{i=1}^{4} Y_{mini\,L2_i} K_{mini\,L3_{1_i}}\right) mod(17)\right) mod(2^4)\right) xor(K_{mini\,L3_2}),$$

где $(x)xor(y)$ – операция «исключающего ИЛИ» над значениями x и y.

Мини-версия блочного симметричного шифра AES подробно рассмотрена в работах [8–13]. Кратко изложим одну из версий мини-шифра и обоснуем ее использование для формирования псевдослучайной подложки в мини-UMAC.

Размер блока открытого текста равен 16 бит, которые обозначим четырьмя шестнадцатеричными числами h_0, h_1, h_2, h_3. Размер ключа также равен 16 бит. Обозначим его как 4 шестнадцатеричных числа k_0, k_1, k_2, k_3.

Шаги шифра применяются к состоянию – массиву 2×2 шестнадцатеричных цифр $\begin{bmatrix} h_0 & h_2 \\ h_1 & h_3 \end{bmatrix}$.

Мини-шифр включает несколько идентичных по структуре раундов (по умолчанию их 4). Перед шифрованием входной блок загружается в состояние, как описано выше, и рассчитываются раундовые ключи. Шифрование имеет общую структуру

$$E(a) = r_4 \circ r_3 \circ r_2 \circ r_1 \circ (a \oplus k_0),$$

где a обозначает состояние, k_0, k_1, k_2, k_3, k_4 – раундовые ключи и

$$r_i(a) = (t \cdot \tilde{\sigma}(S(a))) \oplus k_i,$$

за исключением r_4, где пропущено умножение на t. В конце шифра состояние сгружается в 16-битный блок в таком же порядке, в котором он загружался.

Теперь опишем отдельные компоненты шифра.

SubBytes: операция S есть выборочная таблица, которая применяется к каждой 16-ричной цифре состояния

$$\begin{bmatrix} h_0 & h_2 \\ h_1 & h_3 \end{bmatrix} \xrightarrow{S} \begin{bmatrix} S(h_0) & S(h_2) \\ S(h_1) & S(h_3) \end{bmatrix},$$

где функция S задается следующей таблицей 1.

Таблица 1

Выборочная таблица, реализующая S-блок Baby-Rijndael

x	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
S(x)	a	4	3	b	8	e	2	c	5	7	6	f	0	1	9	d

ShiftRows: операция $\tilde{\sigma}$ просто меняет входы во второй строке состояния

$$\begin{bmatrix} h_0 & h_2 \\ h_1 & h_3 \end{bmatrix} \xrightarrow{\tilde{\sigma}} \begin{bmatrix} h_0 & h_2 \\ h_3 & h_1 \end{bmatrix}.$$

MixColumns: матрица t является циркулянтной 8×8 матрицей, заданной следующей последовательностью бит: [1 0 1 0 0 0 1 1]. Для этого преобразования состояние рассматривается как 8×2 битовая матрица и умножается слева по модулю 2 на t: $a = ta$.

KeySchedule: в начале шифра и в конце каждого раунда состояние побитно складывается (т.е. по модулю 2) с раундовым ключом. Столбцы раундовых ключей определены рекурсивно следующим образом:

$$w_0 = \begin{pmatrix} k_0 \\ k_1 \end{pmatrix}, \; w_1 = \begin{pmatrix} k_2 \\ k_3 \end{pmatrix},$$

$$w_{2i} = w_{2i-2} \oplus S(reverse(w_{2i-2})) \oplus r_i,$$

$$w_{2i+1} = w_{2i-1} \oplus w_{2i}$$

для всех $i = 1,2,3,4$, где $r_i = \begin{pmatrix} 2^{i-1} \\ 0 \end{pmatrix}$, а функция *reverse* заменяет два входа в столбец. Функция *S* та же, что и описанная выше.

Следует заметить, что все сложения выполняются побитно по модулю 2. Наконец, для $i = 1,2,3,4$ раундовый ключ k_i есть матрица, чьи столбцы есть w_{2i} и w_{2i+1}.

Использование рассмотренной уменьшенной модели блочного симметричного шифра AES позволяет провести экспериментальные исследования коллизионных свойств формируемых псевдослучайных подложек по всему множеству секретных ключей. Так, псевдослучайная подложка $Pad_{\min i}$ мини-UMAC формируется посредством шифрования неповторяющегося для каждого информационного сообщения $M_{\min i}$ числа *Nonce*. Результирующее значение $Pad_{\min i}$ имеет длину 16 бит, так же, как и соответствующая длина хеш-кода $Y_{\min i}$.

Мини-версия заключительного преобразования для формирования кодов аутентификации сообщений мини-UMAC состоит в поразрядном суммировании по модулю 2 значений $Y_{\min i}$ и $Pad_{\min i}$: $Tag_{\min i} = Y_{\min i} \oplus Pad_{\min i}$.

Таким образом, масштабирование применяемых преобразований на соответствующих слоях схемы формирования кодов аутентификации сообщений позволяет построить уменьшенную модель UMAC, экспериментально исследовать коллизионные свойства формируемых образов (кодов).

Коэффициент масштабирования при разработке мини-модели UMAC выбран таким образом, чтобы длина формируемых хеш-кодов Y, псевдослучайных подложек Pad и кодов аутентификации сообщений $Tag = Y \oplus Pad$ была равна длине блока мини-версии блочного симметричного шифра AES [8–13], т.е. 16 битам. Выбор такого коэффициента масштабирования позволяет с одной стороны сохранить алгебраическую структуру основных преобразований алгоритма UMAC, в том числе и входящего в его схему алгоритма AES, с другой стороны это дает возможность провести экспериментальные исследования с использованием методов статистической проверки гипотез и математической статистики, рассматривая ограниченный набор элементов Y, Pad и $Tag = Y \oplus Pad$ и соответствующие результаты по оценке числа коллизий как выборку из генеральной совокупности.

Обоснуем методику оценки коллизионных свойств формируемых элементов (обозначим их для простоты $h(x)$), рассмотрим основные условия и ограничения при проведении экспериментов.

Методика оценки коллизионных свойств мини-UMAC. Проведение экспериментальных исследований коллизионных свойств кодов аутентификации сообщений UMAC проведем по соответствующим слоям преобразования:

1. На первом этапе исследуем коллизионные свойства мини-версии универсального хеширования. Для этого необходимо подтвердить в ходе эксперимента теоретические оценки числа возникающих коллизий формируемых хеш-кодов $Y_{\min i}$.

2. На втором этапе проведем экспериментальные исследования коллизионных свойств псевдослучайных подложек $Pad_{\min i}$ на основе анализа свойств уменьшенной модели симметричного шифра. Подобные исследования в доступной литературе не описаны и, по всей видимости, проводятся нами впервые.

3. На третьем этапе проведем экспериментальные исследования коллизионных свойств формируемых с использованием мини-UMAC кодов аутентификации сообщений $Tag_{\min i} = Y_{\min i} \oplus Pad_{\min i}$. Это наиболее важная часть проводимых исследований, поскольку она позволит ответить на вопрос о сохранении свойств универсального хеширования после применения слоя криптографического преобразования информации.

Оценку числа коллизий формируемых элементов будем проводить, ориентируясь на коллизионные свойства универсального хеширования. Собственно говоря, нам требуется подтвердить или опровергнуть гипотезу о сохранении коллизионных свойств универсального хеширования на всех этапах формирования кодов аутентификации сообщений мини-UMAC.

Идея универсального хеширования [13] заключается в определении такого набора элементов конечного множества H хеш-функций $h: A \to B$, $|A| = a$, $|B| = b$ чтобы случайный выбор функции $h \in H$ обеспечивал бы низкую вероятность коллизии, т.е. для любых различных входов x_1 и x_2 вероятность того, что $h(x_1) = h(x_2)$ (вероятность коллизии, столкновения) не должна превосходить некоторой заранее заданной величины ε:

$$P_{кол} = P(h(x_1) = h(x_2)) \le \varepsilon,$$

причем вероятность коллизии может быть рассчитана как

$$P_{кол} = \frac{\delta_H(x_1, x_2)}{|H|},$$

где $\delta_H(x_1, x_2)$ есть количество таких хеш-функций в H, при которых значения $x_1, x_2 \in A$, $x_1 \ne x_2$ вызывают коллизию, т.е. $h(x_1) = h(x_2)$.

Приведем два определения универсального хеширования [5, 13, 14].

1. Пусть $0 < \varepsilon < 1$. H является ε – *универсальным хеш-классом (сокращенно* $\varepsilon - U(H, A, B)$*),* если для двух различных элементов $x_1, x_2 \in A$ существует не больше чем $H \cdot \varepsilon$ функций $f \in H$ таких, что $h(x_1) = h(x_2)$, если $\delta_H(x_1, x_2) \le \varepsilon |H|$ для всех $x_1, x_2 \in A$, $x_1 \ne x_2$.

2. Пусть $0 < \varepsilon < 1$. H является ε – *строго универсальным хеш-классом* (сокращенно $\varepsilon - SU(H, A, B)$), если выполняются следующие условия:

– для каждого $x_1 \in A$ и для каждого $y_1 \in B$,

$$\left|\{h \in H : h(x_1) = y_1\}\right| = |H|/|B|;$$

– для каждого $x_1, x_2 \in A$, $x_1 \ne x_2$ и для каждого $y_1, y_2 \in B$,

$$\left|\{h \in H : h(x_1) = y_1, h(x_2) = y_2\}\right| \le \varepsilon |H|.$$

Определение универсального класса хеш-функций эквивалентно определению такого алгоритма формирования кода аутентификации, при котором число различных правил формирования кода аутентификации (число ключей), при которых существует коллизия (совпадение кодов аутентификации) для двух произвольных входных последовательностей, ограничено. Число таких ключей не может превосходить значение $P_{кол} \cdot |H|$, где $P_{кол}$ – вероятность коллизии, $|H|$ – число всех правил (ключей).

Определение строго универсального класса хеш-функций эквивалентно определению такого алгоритма формирования кодов аутентификации, при котором будут выполняться следующие условия:

1. Число правил формирования кода аутентификации (число ключей), при которых для произвольной входной последовательности значение кода аутентификации не изменяется, ограничено. Число таких ключей не может превосходить значения $|H|/|B|$, где $|H|$ – число всех ключей, $|B|$ – число возможных состояний кода аутентификации.

2. Число правил формирования кода аутентификации (число ключей), при которых для двух произвольных входных последовательностей соответствующие им значения кода аутентификации не изменяются, ограничено. Число таких ключей не может превосходить значения $P_{кол}|H|/|B|$, где $P_{кол}$ – вероятность коллизии, $|H|$ – число всех ключей, $|B|$ – число возможных состояний кода аутентификации.

Вероятность коллизии кодов аутентификации в схеме со строго универсальным хешированием определяется как $P_{кол} \le \varepsilon$.

В основе предлагаемой методики статистического исследования коллизионных свойств формируемых элементов $h(x)$ лежит эмпирическая оценка максимумов числа ключей (правил хеширования) при которых:

1. Для произвольных $x_1, x_2 \in A$, $x_1 \ne x_2$ выполняется равенство

$$h(x_1) = h(x_2). \quad (1)$$

2. Для произвольных $x_1 \in A$ и $y_1 \in B$ выполняется равенство

$$h(x_1) = y_1. \quad (2)$$

3. Для произвольных $x_1, x_2 \in A$, $x_1 \ne x_2$ и $y_1, y_2 \in B$ выполняются равенства

$$h(x_1) = y_1, h(x_2) = y_2. \quad (3)$$

Оценка по первому критерию соответствует проверке выполнимости условия для универсального класса хеш-функций, оценка по второму и третьему критерию – условий для строго универсального класса хеш-функций.

Введем следующие обозначения:

$$n_1(x_1, x_2) = \left|\{h \in H : h(x_1) = h(x_2)\}\right|, \; x_1, x_2 \in A, \; x_1 \ne x_2;$$

$$n_2(x_1, y_1) = \left|\{h \in H : h(x_1) = y_1\}\right|, \; x_1 \in A, \; y_1 \in B;$$

$$n_3(x_1, x_2, y_1, y_2) = \left|\{h \in H : h(x_1) = y_1, h(x_2) = y_2\}\right|, \; x_1, x_2 \in A, \; x_1 \ne x_2, \; y_1, y_2 \in B.$$

Первый показатель $n_1(x_1, x_2)$ характеризует число правил хеширования, при которых для заданных $x_1, x_2 \in A$, $x_1 \ne x_2$ выполняется равенство (1), т.е. число ключей, при которых существует коллизия (совпадение хеш-кодов) для двух входных последовательностей x_1 и x_2.

Второй показатель $n_2(x_1, y_1)$ характеризует число правил хеширования, при которых для заданных $x_1 \in A$, $y_1 \in B$ выполняется равенство (2), т.е. число ключей, при которых для входной последовательности x_1 значение хеш-кода y_1 не изменяется.

Третий показатель $n_3(x_1, x_2, y_1, y_2)$ характеризует число правил хеширования, при которых для заданных $x_1, x_2 \in A$, $x_1 \ne x_2$, $y_1, y_2 \in B$ выполняется равенство (3), т.е. число ключей, при которых для двух входных последовательностей x_1 и x_2 соответствующие им значения хеш-кодов y_1 и y_2 не изменяются.

Поскольку число ключей, при которых могут выполняться равенства (1), (2) и (3), не должно превосходить соответствующих им значений $P_{кол} \cdot |H|$, $|H|/|B|$ и $P_{кол}|H|/|B|$ нас будет интересовать максимальное число таких ключей для каждого из рассматриваемого набора элементов.

Ограничимся изучением статистических характеристик максимумов числа ключей, при которых выполняются равенства (1), (2) и (3), а затем сравним полученные результаты с числом $P_{кол} \cdot H$ (для первого критерия), с числом $|H|/|B|$ (для второго критерия) и числом $P_{кол}|H|/|B|$ (для третьего критерия).

Таким образом, в качестве статистических показателей оценки коллизионных свойств, по которым будем проводить экспериментальные исследования, предлагается использовать:

- математические ожидания $m(n_1)$, $m(n_2)$ и $m(n_3)$ максимумов числа правил хеширования, при которых выполняются равенства (1), (2) и (3), соответственно;

- дисперсии $D(n_1)$, $D(n_2)$ и $D(n_3)$, характеризующие рассеивание значений максимумов числа правил хеширования, при которых выполняются равенства (1), (2) и (3), относительно их математических ожиданий $m(n_1)$, $m(n_2)$ и $m(n_3)$, соответственно.

Оценку коллизионных свойств по приведенным критериям будем производить в среднестатистическом смысле. Другими словами, при постановке эксперимента будем использовать ограниченный набор элементов $x_1, x_2 \in A$, $x_1 \neq x_2$ и соответствующих им хеш-образов $y_1, y_2 \in B$, рассматривая соответствующие результаты как выборку из генеральной совокупности.

Естественной оценкой для математического ожидания m случайной величины X является среднее арифметическое ее наблюдаемых значений X_i (или статистическое среднее) [15]

$$\widetilde{m} = \frac{1}{N}\sum_{i=1}^{N} X_i ,$$

где N – количество реализаций случайной величины X.

Оценка дисперсии случайной величины X определяется выражением

$$\widetilde{D} = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \widetilde{m})^2 .$$

В силу центральной предельной теоремы теории вероятностей при больших значениях количества реализаций N среднее арифметическое будет иметь распределение, близкое к нормальному [15] с математическим ожиданием

$$m[\widetilde{m}] \approx \widetilde{m}$$

и средним квадратическим отклонением

$$\sigma[\widetilde{m}] \approx \frac{\sigma}{\sqrt{N}},$$

где σ – среднее квадратическое отклонение оцениваемого параметра.

При этом вероятность того, что оценка $\widetilde{m}$ отклонится от своего математического ожидания меньше чем на ε (доверительная вероятность) равна [15]

$$P(|\widetilde{m} - m| < \varepsilon) \approx 2\Phi\left(\frac{\varepsilon}{\sigma[\widetilde{m}]}\right), \qquad (4)$$

где $\Phi(x)$ – функция Лапласа, определяется выражением

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}\,dt . \qquad (5)$$

Таким образом, при проведении экспериментальных исследований коллизионных свойств кодов аутентификации сообщений будем использовать методы статистической проверки гипотез и математической статистики. Предлагаемая методика включает следующую последовательность операций.

1. Из генеральной совокупности случайной величины X сформируем выборку объема N: X1, X2, ..., XN:

- для среднестатистической оценки математического ожидания $m(n_1)$ и дисперсии $D(n_1)$ в качестве случайной величины выступает максимум $n_1(x_1, x_2)$ по всем $h(x_1) = h(x_2)$ для заданных x_1 и x_2, следовательно, выборку сформируем отбором из N пар элементов $x_1, x_2 \in A$, $x_1 \neq x_2$;

- для среднестатистической оценки математического ожидания $m(n_2)$ и дисперсии $D(n_2)$ в качестве случайной величины выступает максимум $n_2(x_1, y_1)$ по всем $y_1 = h(x_1)$, следовательно, выборку сформируем отбором из N пар элементов $x_1 \in A$, $y_1 \in B$;

- для среднестатистической оценки математического ожидания $m(n_3)$ и дисперсии $D(n_3)$ в качестве случайной величины выступает максимум $n_3(x_1, x_2, y_1, y_2)$ по всем парам $y_1 = h(x_1)$ и $y_2 = h(x_2)$, следовательно, выборку сформируем отбором из N четверок элементов $x_1, x_2 \in A$, $x_1 \neq x_2$, $y_1, y_2 \in B$.

2. При экспериментальных исследованиях коллизионных свойств хеширования:

- по первому критерию оценим среднее арифметическое $\widetilde{m}(n_1)$ наблюдаемых значений максимумов $n_1(x_1, x_2)$ и дисперсию $\widetilde{D}(n_1)$;

- по второму критерию оценим среднее арифметическое $\widetilde{m}(n_2)$ наблюдаемых значений максимумов $n_2(x_1, y_1)$ и дисперсию $\widetilde{D}(n_2)$;

- по третьему критерию оценим среднее арифметическое $\widetilde{m}(n_3)$ наблюдаемых значений максимумов $n_3(x_1, x_2, y_1, y_2)$ и дисперсию $\widetilde{D}(n_3)$.

3. Обоснуем достоверность полученных среднестатистических оценок. Для этого зафиксируем точность ε и рассчитаем значения функции Лапласа,

которые, в соответствии с выражением (4), дадут соответствующие доверительные вероятности:

$$P(|\tilde{m}(n_1)-m(n_1)|<\varepsilon)\approx 2\Phi\left(\frac{\varepsilon}{\sigma[\tilde{m}(n_1)]}\right),$$

$$\sigma[\tilde{m}(n_1)]\approx\frac{\sqrt{\tilde{D}(n_1)}}{\sqrt{N}};$$

$$P(|\tilde{m}(n_2)-m(n_2)|<\varepsilon)\approx 2\Phi\left(\frac{\varepsilon}{\sigma[\tilde{m}(n_2)]}\right),$$

$$\sigma[\tilde{m}(n_2)]\approx\frac{\sqrt{\tilde{D}(n_2)}}{\sqrt{N}};$$

$$P(|\tilde{m}(n_3)-m(n_3)|<\varepsilon)\approx 2\Phi\left(\frac{\varepsilon}{\sigma[\tilde{m}(n_3)]}\right),$$

$$\sigma[\tilde{m}(n_3)]\approx\frac{\sqrt{\tilde{D}(n_3)}}{\sqrt{N}}.$$

При обратной постановке задачи, т.е. для фиксированной доверительной вероятности P_∂ при объеме выборки N, доверительный интервал определим следующим образом:

$$\tilde{m}-t_\rho\cdot\sigma[\tilde{m}]<m<\tilde{m}+t_\rho\cdot\sigma[\tilde{m}], \qquad (6)$$

где t_ρ – корень уравнения $2\Phi(t_\rho)=P_\partial$.

Другими словами, в этом случае пределы доверительного интервала будут соответствовать заданной доверительной вероятности P_∂, а точность оценок определяется как $\varepsilon=t_\rho\cdot\sigma[\tilde{m}]$.

Для нашего случая при заданной вероятности P_∂ имеем:

$$\tilde{m}(n_1)-t_\rho\cdot\frac{\sqrt{\tilde{D}(n_1)}}{\sqrt{N}}<m(n_1)<\tilde{m}(n_1)+t_\rho\cdot\frac{\sqrt{\tilde{D}(n_1)}}{\sqrt{N}},$$

$$\varepsilon=t_\rho\cdot\frac{\sqrt{\tilde{D}(n_1)}}{\sqrt{N}};$$

$$\tilde{m}(n_2)-t_\rho\cdot\frac{\sqrt{\tilde{D}(n_2)}}{\sqrt{N}}<m(n_2)<\tilde{m}(n_2)+t_\rho\cdot\frac{\sqrt{\tilde{D}(n_2)}}{\sqrt{N}},$$

$$\varepsilon=t_\rho\cdot\frac{\sqrt{\tilde{D}(n_2)}}{\sqrt{N}};$$

$$\tilde{m}(n_3)-t_\rho\cdot\frac{\sqrt{\tilde{D}(n_3)}}{\sqrt{N}}<m(n_3)<\tilde{m}(n_3)+t_\rho\cdot\frac{\sqrt{\tilde{D}(n_3)}}{\sqrt{N}},$$

$$\varepsilon=t_\rho\cdot\frac{\sqrt{\tilde{D}(n_3)}}{\sqrt{N}}.$$

Таким образом, предлагаемая методика, используя уменьшенные модели отдельных слоев преобразований на основе оценки распределения столкновений формируемых образов, позволяет экспериментально исследовать коллизионные свойства кодов аутентификации сообщений.

Выводы

Проведенные исследования показали, что современные многослойные конструкции кодов аутентификации сообщений (UMAC-коды) используют развитый математический аппарат универсального хеширования совместно с блочным симметричным шифрованием.

Для исследования коллизионных свойств кодов аутентификации сообщений UMAC с многослойной конструкцией на основе универсального хеширования и криптографического преобразования предложена методика статистического тестирования. В основе методики лежит использование уменьшенных моделей отдельных слоев преобразований и оценка распределения коллизий (столкновений) формируемых образов (кодов), что позволяет экспериментально исследовать коллизионные свойства кодов аутентификации сообщений мини-UMAC и вырабатывать практические рекомендации по построению эффективных механизмов обеспечения аутентичности и целостности данных на основе полной версии UMAC.

Перспективным направлением является проведение экспериментальных исследований коллизионных свойств мини-UMAC с использованием разработанной методики статистического тестирования, т.е. эмпирическая оценка числа правил хеширования, при которых существует коллизия, а также обоснование на основе опытных данных конкретных предложений по совершенствованию механизмов обеспечения целостности и аутентичности данных в информационных системах.

Список литературы

1. J. Black, S. Halevi, H. Krawczyk, T. Krovetz, and P. Rogaway, "UMAC: Fast and provably secure message authentication", Advances in Cryptology - CRYPTO '99, LNCS vol. 1666. – P. 216-233, Springer-Verlag, 1999.

2. T. Krovetz, and P. Rogaway. Fast universal hashing with small keys and no preprocessing", work in progress, 2000. To be available from http://www.cs.ucdavis.edu/~rogaway/umac

3. T. Krovetz, J. Black, S. Halevi, A. Hevia, H. Krawczyk, and P. Rogaway. UMAC -Message authentication code using universal hashing. IETF Internet Draft, draft-krovetz-umac-00.txt, www.cs.ucdavis.edu/~rogaway/umac, 2000.

4. T. Krovetz. UMAC -Message authentication code using universal hashing. IETF Internet Draft, draft-krovetz-umac-02.txt, www.cs.ucdavis.edu/~rogaway/umac, 2004.

5. Final report of European project number IST-1999-12324, named New European Schemes for Signatures, Integrity, and Encryption, April 19, 2004 – Version 0.15 (beta), Springer-Verlag.

6. T. Krovetz. UMAC - Message authentication code using universal hashing, 2006. To be available from http://www.cs.ucdavis.edu/~rogaway/umac

7. T. Krovetz. Software-Optimized Universal Hashing and Message Authentication. Dissertation submitted in partial satisfaction of the requirements for the degree of doctor of philosophy. University Of California Davis. September 2000. – 269 p.

8. A Description of Baby Rijndael // ISU CprE/Math 533; NTU ST765-U. – 2003.

9. Raphael Chung-Wei Phan, "Mini Advanced Encryption Standard (Mini-AES): A testbed for Cryptanalysis Students", Cryptologia, XXVI(4), October 2002. – P. 283–306.

10. Долгов В.И. Исследование дифференциальных свойств мини-шифров Baby-ADE и Baby-AES / В.И. Долгов, А.А. Кузнецов, Р.В. Сергиенко, О.И. Олешко // Прикладная радиоэлектроника. – Х.: ХНУРЭ, 2009. – Т. 8, № 3. – С. 252-257.

11. Долгов В.И. Подход к криптоанализу современных шифров // Материалы второй международной конференции "Современные информационные системы" / В.И. Долгов, И.В Лисицкая, Р.В. Олейников. Проблемы и тенденции развития:– Харьков-Туапсе, Украина, 2–5 октября. – 2007. – С. 435–436.

12. Сорока Л.С., Кузнецов А.А., Московченко И.В., Исаев С.А. Исследование дифференциальных свойств блочно-симметричных шифров / Л.С. Сорока, А.А. Кузнецов, И.В. Московченко, С.А. Исаев // Системи обробки інформації. – Харків: ХУПС. – 2010. – Вип. 6(87). – С. 286-294.

13. Carter J. L. Universal classes of hash functions / J.L.Carter, M.N.Wegman // Computer and System Scince – 1979. – №18. – P. 143-154.

14. Wegman M. N. New hash functions and their use in authentication and set equality / M. N.Wegman, J. L. Carter / Computer and System Scince – 1981. – № 22 – P. 265–279.

15. Вентцель Е.С. Теория вероятностей / Е.С. Вентцель. – М.: Государственное издательство физико-математической литературы, 1958. – 564 с.

Методика дослідження колізійних властивостей кодів автентифікації повідомлень

О.О. Кузнецов, О.Г. Король, В.В. Босько, С.П. Євсеєв

Розглядаються схема формування кодів автентифікації повідомлень, заснована на використанні універсального гешування (UMAC - Message Authentication Code using Universal Hashing). Пропонується методика статистичного дослідження колізійних властивостей UMAC. В основі запропонованої методики лежить використання зменшеної моделі кодів автентифікації повідомлень (міні-UMAC), побудованої за допомогою масштабування застосовуваних перетворень зі збереженням їх алгебраїчної структури.

Ключові слова: *коди автентифікації повідомлень, колізійні властивості, автентифікація, MAC-код UMAC.*

Methods of collision characteristic research of the message authentification codes

A.A. Kuznetsov, O.G. Korol, V.V. Bosko, S.P. Evseyev

A scheme of the shaping the codes of messages authentification, based on UMAC - Message Authentication Code using Universal Hashing is considered. A method of statistical study of UMAC collision characteristics is offered. Employment of reduced model of message authentification codes (mini-UMAC), built by means of scaling the applied transformations with conservation of their algebraic structure serves as a base for the proposed method.

Keywords: *message authentification codes, collision characteristics, authentification, MAC-code UMAC.*