УДК 004.056

А.А. Кузнецов, Е.П. Колованова, А.А. Пушкарский, З.А. Харченко

Харьковский национальный университет радиоэлектроники, Харьков

КОЛЛИЗИОННЫЕ СВОЙСТВА UMAC

Рассматривается алгоритм формирования кодов аутентификации сообщений UMAC, в основе которого лежит использование универсальных хеширующих функций. Предлагается методика статистического исследования коллизионных свойств формируемых кодов аутентификации сообщений с использованием уменьшенной модели UMAC (mini-UMAC). Исследуются коллизионные свойства кодов аутентификации. Показано, что применение криптографического преобразования (с использованием алгоритма AES) на завершающем этапе UMAC приводит к нарушению свойств универсального хеширования.

Ключевые слова: *хеш-функция, универсальная хеш-функция, алгоритм UMAC, mini-UMAC, коллизия.*

Постановка проблемы в общем виде и анализ литературы

Эффективным механизмом обеспечения целостности и аутентичности информации в современных телекоммуникационных системах и сетях является хеширование информации [1-7], которое применяется как для формирования кодов обнаружения манипуляций (MDC – Manipulation Detection Code) так и для построения кодов аутентификации сообщений (MAC – Message Authentication Code) [5, 7].

Проведенный анализ показал, что наибольшей вычислительной эффективностью обладает отобранный при проведении европейского конкурса NESSIE алгоритм UMAC (Message Authentication Code using Universal Hashing) [5, 7], для формирования кодов аутентификации в котором используются семейства универсальных хеширующих функций [8, 9]. Число коллизий (столкновений) формируемых хеш-образов для каждого введенного ключа универсального хеширования не превышает некоторой заранее заданной величины, а криптостойкость UMAC обеспечивается на уровне выбранного криптоалгоритма (по спецификации рекомендован алгоритм шифрования AES). Однако влияние используемого криптоалгоритма на коллизионные свойства кодов подлинности сообщений UMAC на сегодняшний день не исследовано, обеспечение свойств универсального хеширования в такой многослойной конструкции не обосновано [1-7].

Целью данной работы является исследование коллизионных свойств хеширующих функций алгоритма UMAC, оценка влияния применяемого криптографического преобразования на последнем этапе формирования кодов аутентификации на обеспечение свойств универсального хеширования. Для этого в первой части статьи приводится общая конструкция уменьшенной схемы формирования кодов аутентификации сообщений mini-UMAC, которая, при сохранении математической структуры основных преобразований за счет уменьшения ключевого пространства и пространства аутентификаторов позволяет оценить число возникающих коллизий. Методика статистического исследования коллизионных свойств формируемых кодов аутентификации сообщений с использованием уменьшенной модели UMAC приводиться во второй части статьи. Результаты моделирования и обсуждение полученных данных приводятся в третьей части статьи, по которым делается вывод о нарушении свойств универсального хеширования UMAC.

1. Уменьшенная модель UMAC (mini-UMAC)

Схема формирования кодов аутентификации сообщений UMAC использует в своей структуре несколько слоев преобразования, в том числе блочный симметричных 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 в целом, т.е. на коллизионные свойства кодов аутентификации сообщений уменьшенной модели UMAC. Схема формирования кодов UMAC состоит из следующих слоев:

– трехуровневое универсальное хеширование для формирования хеш-кодов

$$\mathrm{Y} = \mathrm{Hash}(\mathrm{K}, \mathrm{M}, \mathrm{Taglen});$$

– криптографическое преобразования с использованием блочного симметричного шифра для формирования псевдослучайной подложки

$$\mathrm{Pad} = \mathrm{PDF}(\mathrm{K}, \mathrm{Nonce}, \mathrm{Taglen});$$

– 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.

1.1. Мини-версия трехуровневого универсального хеширования

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

Соответствующая длина хеш-кода Y_{mini} уменьшенной модели первого слоя будет кратна 4 битам, его значение сформируем посредством объединения (конкатенации) 4-х последовательностей Y_{miniL3_i}

$$Y_{mini} = Y_{miniL3_1} \| Y_{miniL3_2} \| Y_{miniL3_3} \| Y_{miniL3_4} ,$$

где Y_{miniL3_i} - результат многоуровневого хеширования сообщения уменьшенной модели первого слоя mini-UMAC.

Рассмотрим процесс формирования одного блока Y_{miniL3_i} (второй уровень хеширования в уменьшенной модели выполнять не будем, так как при данной размерности он не несет смысловой нагрузки):

$$Y_{miniL3_i} = Y_{miniL3} = Hash_{miniL3} \times$$

$$\times\left(K_{miniL3_1}, K_{miniL3_2}, Hash_{miniL1}\left(K_{miniL1}, M_{mini}\right)\right),$$

где K_{miniL1}, K_{miniL3_1}, K_{miniL3_2} - ключевые последовательности mini-UMAC, $Hash_{miniL1}$ и $Hash_{miniL3}$ - уменьшенные версии хеширования первого и третьего уровней соответственно.

На первом уровне массив-строка M_{mini} размерности 32 бита преобразуется функцией $NH(K_{L1}, M_i)$. Эта строка и является результатом хеширования первого уровня:

$$Y_{miniL1} = NH_{mini}\left(K_{miniL1}, M_{mini}\right).$$

Значение функции $NH_{mini}\left(K_{miniL1}, M_{mini}\right)$ вычисляется по следующему правилу. Информационный блок M_{mini} 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: $K_{miniL1} = K_{miniLl_1} \| K_{miniLl_2} \| .. \| K_{miniLl_8}$, после чего (принимая начальное состояние $Hash_{L1} = 0$) выполняются следующие операции:

$$Hash_{miniL1} = Hash_{miniL1} +_8 ((M_{mini_0} +_4 K_{miniLl_0}) \times_8$$
$$\times_8 (M_{mini_4} +_4 K_{miniLl_4}));$$

$$Hash_{miniL1} = Hash_{miniL1} +_8 ((M_{mini_1} +_4 K_{miniLl_1}) \times_8$$
$$\times_8 (M_{mini_5} +_4 K_{miniLl_5}));$$

$$Hash_{miniL1} = Hash_{miniL1} +_8 ((M_{mini_2} +_4 K_{miniLl_2}) \times_8$$
$$\times_8 (M_{mini_6} +_4 K_{miniLl_6}));$$

$$Hash_{miniL1} = Hash_{miniL1} +_8 ((M_{mini_3} +_4 K_{miniLl_3}) \times_8$$
$$\times_8 (M_{mini_7} +_4 K_{miniLl_7}));$$

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

В результате вычислений формируется восьмибитное значение $Y_{miniL1} = Hash_{miniL1}$.

Третий уровень хеширования преобразует поданные на его вход восьмибитные данные Y_{miniL1} в хеш-код Y_{miniL3} длины 4 бита. В качестве ключевых последовательностей выступают K_{miniL3_1} и K_{miniL3_2} длины 16 и 4 бита соответственно.

Хешируемые данные $Hash_{miniL1}$ и ключевая последовательность K_{miniL3_1} равномерно разбиваются на четыре блока, каждый из которых представляется как целое число Y_{miniL2_i} и $K_{miniL3_{1_i}}$, $i = 1,2,...,4$.

Хеш-значение Y_{miniL3} вычисляется следующим образом:

$$Y_{miniL3} = (K_{miniL3_2}) \text{xor}$$

$$\text{xor}\left(\left(\left(\sum_{i=1}^{4} Y_{miniL2_i} K_{miniL3_{1i}}\right) \bmod(17)\right) \bmod(2^4)\right),$$

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

1.2. Мини-версия блочного симметричного шифра AES

Мини-версия AES для формирования псевдослучайной подложки подробно рассмотрена в работах [10 - 14]. Наиболее простой в реализации есть мини-версия шифра AES (Baby-Rijndael), которая предложена К. Бергманом [10]. Детальнее ознакомиться с уменьшенную моделью шифра AES можно в вышеперечисленных источниках.

Размер блока открытого текста равен 16 бит, которые обозначим четырьмя шестнадцатеричными числами h_0, h_1, h_2, h_3. Отметим, что h_0 состоит из первых четырех бит входного потока. Однако когда h_0 рассматривается как шестнадцатеричная цифра, первый бит рассматривается как бит высшего порядка. Например, входной блок 1000 1100 0111 0001 будет представлен $h_0 = 8$, $h_1 = c$, $h_2 = 7$, $h_3 = 1$.

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 2Ч2 шестнадцатеричных цифр. Однако для рассматриваемой ниже операции $\tilde{\sigma}$ состояние будет

представлено как массив 8Ч2 бит, т.е. каждая шестнадцатеричная цифра будет рассматривается как столбец 4 бит с битом высшего порядка сверху.

Входной блок загружается в состояние отображением h_0, h_1, h_2, h_3 в $\begin{bmatrix} h_0 & h_2 \\ h_1 & h_3 \end{bmatrix}$. Например, входной блок 1000 1100 0111 0001 загружается как

$\begin{bmatrix} 8 & 7 \\ c & 1 \end{bmatrix}$, где матрица 8Ч2 – $\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$.

Baby-Rijndael 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 табл. 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 следующей 8Ч8 матрицей бит:

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

Для этого преобразования состояние рассматривается как 8Ч2 битовая матрица. Состояние умножается слева на t, используя матричное умножение по модулю 2: $a = ta$.

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 таким образом:

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

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

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

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_{mini} mini-UMAC формируется посредством шифрования неповторяющегося для каждого информационного сообщения M_{mini} числа Nonce. Результирующее значение Pad_{mini} имеет длину 16 бит, как и соответствующая длина хеш-кода Y_{mini}.

1.3. Мини-версия заключительного преобразования

Формирование кодов аутентификации сообщений mini-UMAC состоит в поразрядном суммировании по модулю 2 значений Y_{mini} и Pad_{mini}:

$$Tag_{mini} = Y_{mini} \oplus Pad_{mini}.$$

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 [10], 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 статистического оценивания коллизионных свойств формируемых элементов (обозначим их для простоты $h(x)$), рассмотрим основные условия и ограничения при проведении экспериментов.

2. Методика статистического исследования коллизионных свойств

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_{mini}.

2. На втором этапе проведем экспериментальные исследования коллизионных свойств псевдослучайных подложек Pad_{mini} на основе анализа свойств уменьшенной модели шифра Baby-Rijndael. 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 mini-UMAC кодов аутентификации сообщений $Tag_{mini} = Y_{mini} \oplus Pad_{mini}$. 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 mini-UMAC.

Идея универсального хеширования 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 некоторую заранее заданную величину 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 [8, 9].

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 $y_1 \in B$,
$|\{h \in H : h(x_1) = y_1\}| = |H|/|B|$;
- для каждого $x_1, x_2 \in A$, $x_1 \ne x_2$ и для каждого $y_1, y_2 \in B$,
$|\{h \in H : h(x_1) = y_1, h(x_2) = y_2\}| \le \varepsilon |H|$.

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 правила:

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 соот-

ветствующие им значения кода аутентификации не изменяются, ограничено. Число таких ключей не может превосходить значения $P_{кол}|H|$, 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 $x_1, x_2 \in A$, $x_1 \neq x_2$ выполняется равенство

$$h(x_1) = h(x_2); \qquad (2.1)$$

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

$$h(x_1) = y_1; \qquad (2.2)$$

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

$$h(x_1) = y_1, h(x_2) = y_2. \qquad (2.3)$$

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:

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

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

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

Первый показатель $n_1(x_1, x_2)$ характеризует число правил хеширования, при которых для заданных $x_1, x_2 \in A$, $x_1 \neq x_2$ выполняется равенство (2.1), 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 (2.2), 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 $x_1, x_2 \in A$, $x_1 \neq x_2$, $y_1, y_2 \in B$ выполняется равенство (2.3), 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 (2.1), (2.2) и (2.3), не должно превосходить соответствующих им значений $P_{кол} \cdot |H|$, $|H|/|B|$ и $P_{кол}|H|/|B|$ проведем оценку максимального числа таких ключей для каждого из рассматриваемого набора элементов.

Ограничимся изучением статистических характеристик максимумов этих величин, 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 предлагается использовать: математические ожидания $m(n_1)$, $m(n_2)$ и $m(n_3)$ максимумов числа правил хеширования, при которых выполняются равенства (2.1), (2.2) и (2.3), соответственно; дисперсии $D(n_1)$, $D(n_2)$ и $D(n_3)$, характеризующие рассеивание значений числа правил хеширования, при которых выполняются равенства (2.1), (2.2) и (2.3), 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. Т.е. 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]

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

где N – количество реализаций случайной величины X.

Оценка дисперсии случайной величины *X* определяется выражением

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

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

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

и средним квадратическим отклонением

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

где σ – среднее квадратическое отклонение оцениваемого параметра.

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

$$P(|\tilde{m} - m| < \varepsilon) \approx 2\Phi\left(\frac{\varepsilon}{\sigma[\tilde{m}]}\right), \qquad (2.4)$$

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

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

Проведение экспериментальных исследований организовано следующим образом.

1. Из генеральной совокупности случайной величины *X* сформируем выборку:

- для среднестатистической оценки $m(n_1)$ и $D(n_1)$ в качестве случайной величины выступает максимум $n_1(x_1, x_2)$ при которых выполняется равенство $h(x_1) = h(x_2)$, следовательно, выборку объема N: X_1, X_2, ..., X_N сформируем отбором N множеств, в каждом из которых содержится M пар элементов $x_1, x_2 \in A$, $x_1 \neq x_2$ и оценивается $n_1(x_1, x_2)$, т.е. общий объем формируемых пар элементов $x_1, x_2 \in A$, $x_1 \neq x_2$ составит NM;

- для среднестатистической оценки $m(n_2)$ и $D(n_2)$ в качестве случайной величины выступает максимум $n_2(x_1, y_1)$ при которых выполняется равенство $y_1 = h(x_1)$, следовательно, выборку сформируем отбором N множеств, в каждом из которых содержится M пар элементов $x_1 \in A$, $y_1 \in B$ и оценивается $n_2(x_1, y_1)$. Общий объем формируемых пар элементов $x_1 \in A$, $y_1 \in B$ составит NM;

- для среднестатистической оценки $m(n_3)$ и $D(n_3)$ в качестве случайной величины выступает максимум $n_3(x_1, x_2, y_1, y_2)$ при которых выполняются равенства $y_1 = h(x_1)$ и $y_2 = h(x_2)$, следовательно, выборку сформируем отбором N множеств, в каждом из которых содержится M четверок элементов $x_1, x_2 \in A$, $x_1 \neq x_2$, $y_1, y_2 \in B$ и оценивается $n_3(x_1, x_2, y_1, y_2)$, общий объем формируемых четверок составит NM.

2. При экспериментальных исследованиях коллизионных свойств хеширования будем оценивать среднее арифметическое $\tilde{m}(n_i)$ наблюдаемых значений максимумов n_i и дисперсию $\tilde{D}(n_i)$, $i = 1,2,3$.

3. Обоснуем достоверность полученных среднестатистических оценок. Зафиксируем точность ε и рассчитаем значения функции Лапласа, которые, в соответствии с выражением (2.4), дадут соответствующие доверительные вероятности:

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

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

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

$$\tilde{m}(n_i) - t_\rho \cdot \sigma[\tilde{m}(n_i)] < m(n_i) < \tilde{m}(n_i) + t_\rho \cdot \sigma[\tilde{m}(n_i)], \quad (2.6)$$

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

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.

3. Результаты моделирования и их обсуждение

С использованием разработанной уменьшенной модели UMAC (mini-UMAC) и методики статистического исследования коллизионных свойств кодов аутентификации сообщений проведем экспериментальную оценку распределения числа столкновений (коллизий) формируемых образов.

Поскольку в рассмотренной выше схеме UMAC на первом слое (при формировании хеш-кода Y_{mini}) используются семейства универсальных хеширующих функций, подробно исследуемые в работах [1-7], статистические исследования проведем только на втором слое (при формировании псевдослучайной подложки Pad_{mini}) и на заключительном этапе формирования кодов аутентификации (после выполнения суммирования $Tag_{mini} = Y_{mini} \oplus Pad_{mini}$). Именно на этих этапах, по нашему предположению и нарушаются свойства универсальности формируемых кодов аутентификации.

При проведении статистических исследований коллизионных свойств формируемых значений Pad_{mini} и Tag_{mini} для каждого эксперимента оценивались математические ожидания $m(n_1)$, $m(n_2)$ и $m(n_3)$, дисперсии $D(n_1)$, $D(n_2)$ и $D(n_3)$, а также для фиксированной точности $\varepsilon = 0{,}1$ рассчитывались соответствующие доверительные вероятности $P(|\tilde{m}(n_i) - m(n_i)| < \varepsilon)$. Исследования проводились над выборкой, объема $N = 100$, для формирования каждого элемента выборки рассчитывался максимум по множеству из $M = 1000$ кортежей элементов. Таким образом, общий объем формируемых наборов составил $NM = 10^5$.

Полученные результаты экспериментальных исследований сведены в табл. 2.

Таблица 2

Результаты экспериментальных исследований

	mini-AES, Pad_{mini}	mini-UMAC, Tag_{mini}
$\tilde{m}(n_1)$	-	4,19
$\tilde{D}(n_1)$	-	0,6
$P_д = P(\lvert\tilde{m}(n_1) - m(n_1)\rvert < \varepsilon)$	-	0,99
$\tilde{m}(n_2)$	4,26	4,23
$\tilde{D}(n_2)$	0,6	0,65
$P_д = P(\lvert\tilde{m}(n_2) - m(n_2)\rvert < \varepsilon)$	0,99	0,99
$\tilde{m}(n_3)$	0,001	0,002
$\tilde{D}(n_3)$	0,001	0,002
$P_д = P(\lvert\tilde{m}(n_3) - m(n_3)\rvert < \varepsilon)$	0,99	0,99

При исследовании коллизионных свойств кодов аутентификации, сформированных с использованием мини-версии шифра AES число ключей, при которых выполняется равенство $h(x_1) = h(x_2)$, при всех испытаниях равнялось нулю, т.е. $n_1(x_1, x_2) = 0$ во всех $N = 100$ опытах. Этот результат объясняется следующим свойством. Шифр AES (как и его мини-версия), реализует биективное отображение множества открытых текстов в множество шифрограмм, т.е. для фиксированного ключа формируемые шифр-тексты, соответствующие различным открытым текстам, будут различны. Проводимые экспериментальные исследования по первому введенному критерию как раз и состояли в подсчете числа ключей, при которых наблюдается столкновение (коллизия) двух шифр-тектов, соответствующих двум различным открытым текстам, что невозможно по определению биективного шифра. В связи с этим статистические данные по первому критерию для мини-версии шифра AES в табл. 2 не приведены, как не информативные.

Анализ приведенных в табл. 2 данных позволяет утверждать об адекватности полученных результатов и соответствии их статистическим свойствам всей генеральной совокупности данных. Для фиксированной точности ε = 0,1 получены высокие значения доверительной вероятности. Сравним полученные результаты среднестатистических оценок $m(n_1)$, $m(n_2)$ и $m(n_3)$ с теоретическим оценками.

Рассмотрим *первый критерий*, по которому оценивается число правил хеширования, при которых существует коллизия (совпадение кодов аутентификации) для двух произвольных входных последовательностей. В соответствии с теоретическими оценками эта величина ограничена сверху числом $P_{кол} \cdot \lvert H\rvert$. Конкретизируем эту (теоретическую) оценку кодов аутентификации, сформированных с использованием mini-AES и mini-UMAC.

Мощность ключевого множества для mini-AES и mini-UMAC составляет $\lvert H\rvert = 2^{16}$, мощность множества формируемых кодов аутентификации также составляет $\lvert B\rvert = 2^{16}$. Если использовать верхнюю оценку вероятности коллизий как обратную величину мощности формируемых кодов аутентификации $P_{кол} = 2^{-16}$ получим $n_1(x_1, x_2) \le P_{кол} \cdot \lvert H\rvert = 1$. Для мини-версии шифра AES это условие выполняется (обосновывается биективностью шифрующего преобразования), однако коллизионные свойства mini-UMAC существенно уступают этой верхней теоретической оценке. Фактически, число коллизий выше теоретической границы более чем в четыре раза и это положение подтверждено с высокой доверительной вероятностью

$$P_д = P(\lvert\tilde{m}(n_1) - m(n_1)\rvert < 0{,}1) > 0{,}99 .$$

Рассмотрим *второй критерий*, по которому оценивается число правил хеширования, при которых для произвольной входной последовательности значение кода аутентификации не изменяется. В соответствии с теоретическими оценками эта величина для кодов аутентификации, сформированных с использованием mini-AES и mini-UMAC, ограничена сверху числом $\lvert H\rvert / \lvert B\rvert = 1$. Полученные экспериментальные результаты свидетельствуют, что коллизионные свойства кодов аутентификации, сформированных с использованием mini-AES и mini-UMAC, не удовлетворяют второму критерию, число ключей, при которых для произвольной входной последовательности значение кода аутентификации не изменяется в несколько раз превышает теоретическую оценку для универсального хеширования.

В соответствии с *третьим критерием* оценивается число правил хеширования, при которых для двух произвольных входных последовательностей соответствующие им значения кода аутентификации не изменяются. Теоретическая оценка этой величины для универсального хеширования ограничена сверху числом $P_{кол}\lvert H\rvert$, что при $P_{кол} = 2^{-16}$ дает $n_3(x_1, x_2, y_1, y_2) \le$ $\le P_{кол} \cdot \lvert H\rvert = 1$. Значения, приведенные в таблице 2, свидетельствуют о том, что коллизионные свойства кодов аутентификации, сформированных как с использованием mini-AES, так и с использованием mini-UMAC удовлетворяют третьему критерию.

Выводы

Таким образом, из полученных результатов статистических исследований коллизионных свойств кодов аутентификации сообщений, сформированных с использованием mini-AES и mini-UMAC, можно сделать важные в прикладном отношении выводы:

- первый слой формирования кодов аутентификации сообщений удовлетворяет свойствам универ-

сального хеширования, вероятность коллизии не превосходит наперед заданной величины. Это объясняется свойствами применяемых функций хеширования, которые строятся с использованием специального математического аппарата NH и полиномиального хеширования, а также модулярных преобразований. В тоже время используемые функции хеширования не являются строго универсальными, т.к. не выполняются ограничения числу правил хеширования, при которых для произвольных входных последовательностей соответствующие им значения кода аутентификации не изменяются;

- второй слой формирования кодов аутентификации сообщений удовлетворяет свойствам универсального хеширования. Это объясняется тем, что шифрование неповторяющегося для всех информационных сообщений значения Nonce приводит к формированию множества уникальных для всех информационных сообщений псевдослучайных подложек Pad , в результате чего коллизии отсутствуют по определению. В тоже время, второй слой формирования кодов аутентификации сообщений не удовлетворяет свойствам строго универсального хеширования;

- третий слой формирования кодов аутентификации сообщений не удовлетворяет свойствам универсального хеширования. Это объясняется тем, что схема с простым суммированием по модулю два двух результатов универсального хеширования не сохраняет их коллизионные свойства.

Перспективным направлением является исследование возможностей по совершенствованию рассмотренной схемы формирования кодов аутентификации сообщений с обеспечением свойств универсального хеширования, обоснование предложений по обеспечению высоких коллизионных свойств в усовершенствованной схеме UMAC.

Список литературы

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

2. Fast universal hashing with small keys and no preprocessing / T. Krovetz, P. Rogaway. – Work in progress, 2000. http://www.cs.ucdavis.edu/~rogaway/umac

3. UMAC -Message authentication code using universal hashing. IETF Internet Draft / T. Krovetz, J. Black, S. Halevi, A. Hevia. –. 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. – 269p.

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

9. 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.

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

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

12. Исследование дифференциальных свойств миниширов 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.

13. Долгов В.И. Подход к криптоанализу современных шифров / В.И. Долгов, И.В. Лисицкая, Р.В. Олейников // Мат. 2-й межд. конференции "Современные информационные системы. Проблемы и тенденции развития". – Харьков-Туапсе, Украина, 2–5 октября. – 2007. – С. 435-436.

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

15. Вентцель Е.С. Теория вероятностей / Е.С. Вентцель . – М.: Гос. изд-во физ.-мат. лит., 1958 – 564 с.

Поступила в редакцию 5.04.2013

Рецензент: д-р техн. наук проф. И.Д. Горбенко, Харьковский национальный университет радиоэлектроники, Харьков.

ДОСЛІДЖЕННЯ КОЛІЗІЙНИХ ВЛАСТИВОСТЕЙ КОДІВ АВТЕНТИФІКАЦІЇ ПОВІДОМЛЕНЬ UMAC

О.О. Кузнецов, Є.П. Колованова, А.О. Пушкарський, З.О. Харченко

Розглядається алгоритм формування кодів автентифікації повідомлень UMAC, в основі якого лежить використання універсальних функцій гешування. Пропонується методика статистичного дослідження колізійних властивостей формованих кодів аутентифікації повідомлень з використанням зменшеної моделі UMAC (mini-UMAC). Досліджуються колізійні властивості кодів автентифікації. Показано, що застосування криптографічного перетворення (з використанням алгоритму AES) на завершальному етапі UMAC призводить до порушення властивостей універсального хешування.

***Ключові слова**: геш-функція, універсальна геш-функція, алгоритм UMAC, mini-UMAC, колізія.*

RESEARCH OF COLLISION PROPERTIES OF MESSAGE AUTHENTICATION CODES UMAC

A.A. Kuznecov, E.P. Kolovanova, A.A. Pushkarskiy, Z.A. Kharchenko

An algorithm for the formation of a message authentication code UMAC, which is based on the use of universal hash-functions. The methods of statistical analysis of collision properties of the generated message authentication codes using the reduced model UMAC (mini-UMAC). Researched collision properties of authentication codes. It is shown that the use of a cryptographic transformation (using an algorithm AES) at the final stage UMAC leads to disruption of the properties of universal hashing.

Keywords: *hash-function, a universal hash-function algorithm UMAC, mini-UMAC, collision.*