УДК 681.3.06:519.248.681

А.Д. ТЕВЯШЕВ, д-р техн. наук, Ю.И. ГОРБЕНКО

ОЦЕНКА ОПАСНОСТИ КРИПТОАНАЛИТИЧЕСКИХ АТАК МЕТОДОМ СОЗДАНИЯ КОЛЛИЗИЙ

Введение

В системах криптографической защиты информации широкое применение находят криптографические преобразования типа криптографической хэш-функции, однонаправленной хэш-функции, цифровой подписи и др. [1-3]. К таким преобразованиям предъявляются жесткие требования по стойкости к созданию коллизий [1]. Под стойкостью к коллизиям понимают вычислительную невозможность нахождения двух сообщений M_i и M_j, для которых

$$H(M_i)=H(M_j), \tag{1}$$

где H есть соответствующее преобразование. В ряде источников [2-3] приводятся оценки вероятности создания коллизий, причем считается, что для этого необходимо выполнить не менее $\sqrt{n}$ экспериментов из общего числа n возможных значений. Анализ состояния вопроса оценки опасности создания коллизии показывает, что актуальными являются по крайней мере две следующие задачи, требующие своего решения.

Задача 1. Пусть имеется некоторая функция преобразования H информации M

$$h=H(M), \tag{2}$$

где M есть информация произвольной длины l_M, причем h может принимать $n=2^m$ значений независимо от длины l_M. Необходимо определить число k случайных сообщений M_i, которые необходимо подать на вход преобразователя H, чтобы с вероятностью P_3 состоялось хотя бы одно совпадение вида (1), т.е. состоялась коллизия.

Задача 2. Пусть на выходе преобразователя H из полного множества значений $n=2^m$ формируются k случайных значений функции преобразования (2), причем $k \le n$ и k подчиняются равновероятному закону распределения. Пусть выполнено z экспериментов, в каждом из которых получено k значений h. Обозначим реализации двух экспериментов соответственно как X и Y, причем $X=(x_1,x_2,\ldots,x_k)$ и $Y=(y_1,y_2,\ldots,y_k)$. Необходимо найти вероятность $P(n,k)$ того, что эти множества содержат в себе хотя бы по одному элементу x_i и y_j, такие, что $x_i=y_j$.

Целью настоящей статьи и является решение задач 1 и 2 в общем виде и обсуждение полученных результатов.

1. Оценка вероятности появления коллизий (задача 1)

Проведенный анализ показал [1], что задача 1 может быть решена с использованием "парадокса" о дне рождения, но при подробном рассмотрении выясняется, что она носит более общий характер. В нашем случае имеется выборка из k значений целочисленной случайной величины с равновероятным законом распределения, причем она может принимать значения от 1 до $n=2^m$, а $k \le n$. При этих условиях необходимо найти вероятность $P(n,k)$ того, что среди значений $H(M)$ выборки по крайней мере две совпадают, т.е.

$$H(M_i)=H(M_j).$$

Для решения задачи 1 найдем вероятность того, что в группе из k событий не состоится коллизия, т.е. соотношение (1) не выполнится ни разу. Обозначим эту вероятность как $R(n,k)$. Ясно, что $P(n,k)$ и $R(n,k)$ составляют полную группу событий, т.е.

$$P(n,k)+R(n,k)=1$$

и

$$P(n,k)=1-R(n,k). \qquad (3)$$

Далее найдем общее число N различных способов, которыми можно получить k значений без повторений. Для первого элемента мы имеем n значений без повторений, для второго n-1, третьего – n-2 и т.д., для k-го (n-k+1). Поэтому общее число способов, при которых совпадений вида (1) нет, равно

$$N = n\cdot(n-1)(n-2)..(n-k+1) = \frac{n!}{(n-k)!} \qquad (4)$$

Поскольку при каждом из событий с одинаковой вероятностью может происходить каждое из n событий, то общее число событий N_Σ можно оценить как

$$N_\Sigma = n^k. \qquad (5)$$

Далее, вероятность отсутствия совпадений можно оценить отношением числа вариантов без совпадений (4) к общему числу вариантов (5), т.е.

$$R(n,k) = \frac{n!}{(n-k)!}/n^k = \frac{n!}{(n-k)!\,n^k}. \qquad (6)$$

Используя соотношение (3), имеем

$$P(n,k) = 1-\frac{n!}{(n-k)!\,n^k}. \qquad (7)$$

Выражение (7) может быть использовано для оценки соответствующей вероятности, однако предпочтительнее получить общее решение уравнения (7), например, для значения k.

Для этого представим $P(n,k)$ в виде:

$$\begin{aligned} P(n,k) &= 1-\frac{n(n-1)...(n-k+1)}{n^k} = \\ &= 1-\left[\frac{n-1}{n}\cdot\frac{n-2}{n}\cdot...\cdot\frac{n-k+1}{n}\right] = \\ &= 1-\left[\left(1-\frac{1}{n}\right)\cdot\left(1-\frac{2}{n}\right)\cdot...\cdot\left(1-\frac{k-1}{n}\right)\right] \end{aligned} \qquad (8)$$

Далее воспользуемся тем, что для всех $x \ge 0$ [1]

$$(1-x) \le e^{-x}. \qquad (9)$$

При малых значениях x (например, $x \le 0{,}1$) можно считать, что

$$(1-x) \approx e^{-x}. \qquad (10)$$

С учетом этого преобразуем выражение (8), подставив значение (10). В результате получим

$$P(n,k)=1-\left(e^{-1/n}\cdot e^{-2/n}\cdot\ldots\cdot e^{-\frac{k-1}{n}}\right)=1-e^{-\left(\frac{1}{n}+\frac{2}{n}+\ldots+\frac{k-1}{n}\right)}=1-e^{-k(k-1)\,2n}. \quad (11)$$

Обозначим $P(n,k)=P_з$, т.е. вероятностью, с которой должна быть коллизия. В результате имеем

$$P_з=1-e^{-k(k-1)/2n}$$

или

$$1-P_з=e^{-k(k-1)/2n}. \quad (12)$$

Прологарифмировав (12), имеем

$$\ln(1-P_з)=-k(k-1)/2n. \quad (13)$$

Преобразуя (13), имеем

$$\frac{k(k-1)}{2n}=-\ln(1-P_з).$$

или

$$k(k-1)=-2n\ln(1-P_з).$$

В конечном виде получаем

$$k^2-k+2n\ln(1-P_з)=0. \quad (14)$$

Таким образом, получено уравнение, в котором связаны три величины – число событий k, общее число событий и вероятность $P(n,k)$, с которой должна осуществляться коллизия. Задаваясь соответствующими значениями $P_з$ и n можно получить точное решение.

Пусть $P_з$=0,5 , тогда из (14) получаем

$$k^2-k+2n\ln 0.5=k^2-k-2n\ln 2=0. \quad (15)$$

При $n=2^m$ уравнение (15) имеет вид

$$k^2-k-2^{m+1}\ln 2=0. \quad (16)$$

Дадим оценку значения k, учитывая, что $k^2>>k$. Из (14) получаем

$$k^2=-2n\ln(1-P_з). \quad (17)$$

При $P_З$=0,5 имеем

$$k^2=-2n\ln(1-0,5)=2n\ln 2$$

и

$$k=\sqrt{2n\ln 2}\approx 1,41\sqrt{n}. \quad (18)$$

При произвольном значении $P_з$ из (17) получим

$$k=\sqrt{2\ln\left(\frac{1}{1-P_з}\right)\cdot n}=1,41\sqrt{\ln\left(\frac{1}{1-P_з}\right)\cdot n}. \quad (19)$$

Соотношение (19) позволяет оценить число преобразований (экспериментов) $H(M)$, которые необходимо выполнить для осуществления коллизии с вероятностью $P_з$.

Соотношение (19) позволяет оценить число преобразований (экспериментов) $H(M)$, которые необходимо выполнить для осуществления коллизии с вероятностью $P_з$.

Сравнивая полученные для k значения, например, (18) и (19) с оценкой, которая приводится в ряде источников [2-3]

$$k = \sqrt{n}, \tag{20}$$

можно оценить степень приближенности и возможность ее применения.

Пример 1. Пусть в качестве H используется хэш-функция SHA-1, в которой $n=2^{160}$, и пусть $P_з^{'} = 0{,}5$ и $P_з^{''} = 0{,}99$. Используя выражение (19), имеем

$$k_{0,5} = 1{,}41\sqrt{n} = 1{,}41\sqrt{2^{160}} = 1{,}41 \cdot 2^{80} \approx 1{,}7 \cdot 10^{24};$$

$$k_{0,99} = 1{,}41\sqrt{\ln\frac{1}{1-0{,}99} \cdot 2^{160}} = 2^{80} \approx 3 \cdot 10^{24}.$$

В случае (20) получаем

$$k = \sqrt{n} = \sqrt{2^{160}} = 2^{80} \approx 1{,}2 \cdot 10^{24}.$$

Из примера видна погрешность, которую дает оценка (20), широко используемая для анализа коллизий.

2. Решение задачи 2 анализа коллизий

Рассмотрим решение задачи 2, опираясь на результаты задачи 1. Вторая задача возникает при рассмотрении и выборе способов создания коллизий [1]. Например, в нашей постановке эта задача имеет смысл, если рассмотрение сразу двух или большего числа множеств из k-реализаций позволяет ускорить процесс создания коллизий или извлечь полезную для криптоаналитика информацию.

В нашем случае событие $x_i=y_j$ может состояться с вероятностью $\frac{1}{n}$. Поэтому вероятность того, что $x_i \neq y_j$

$$Q(x_i \neq y_j) = 1 - \frac{1}{n}. \tag{21}$$

Если Y включает в себя k событий, то вероятность того, что все значения $y_1, y_2, \ldots, y_k$ не совпадут с x_i, может быть вычислена как

$$Q(x_i \neq Y) = \left(1 - \frac{1}{n}\right)^k. \tag{22}$$

Вероятность того, что хотя бы одно значение Y совпадет с x_i, есть

$$R(x_i \in Y) = 1 - \left(1 - \frac{1}{n}\right)^k. \tag{23}$$

Пусть все элементы X разные. Это справедливо, так как $k<<n$, например, $k=\sqrt{n}$. Тогда вероятность того, что

$$R(x_i \notin Y) = \left(1 - \frac{1}{n}\right)^k$$

и

Далее, вероятность того, что хотя бы одно событие из X и Y совпадет, есть

$$R(x_i = y_j) = 1 - \left(1 - \frac{1}{n}\right)^{k^2}. \quad (25)$$

Обозначим $x = \frac{1}{n} << 1$ и воспользуемся соотношением (10). В результате получим

$$R(n,k) = 1 - \left(1 - \frac{1}{n}\right)^{k^2} = \left(1 - \left(e^{-1/n}\right)^{k^2}\right) = 1 - e^{-\frac{k^2}{n}}. \quad (26)$$

Таким образом, вероятность того, что в двух множествах X и Y хотя бы по одному элементу совпадают,

$$P(n,k) = P_з = 1 - e^{-k^2/n}. \quad (27)$$

Преобразуя (27), получим

$$e^{-k^2/n} = 1 - P_з. \quad (28)$$

Логарифмируя (28), имеем

$$-\frac{k^2}{n} = \ln(1 - P_з)$$

и далее

$$k^2 = -n \ln\left(\frac{1}{1 - P_з}\right)^{-1} = n \ln\left(\frac{1}{1 - P_з}\right).$$

В заключение получим

$$k = \sqrt{n \ln\left(\frac{1}{1 - P_з}\right)}. \quad (29)$$

Рассмотрим частые случаи.
Пусть $P_з$=0,5, тогда

$$k = \sqrt{n \ln \frac{1}{1 - 0{,}5}} = \sqrt{n \ln 2} = 0{,}83\sqrt{n}. \quad (30)$$

При $P_з$=0,99 получим

$$k = \sqrt{n \ln \frac{1}{1 - 0{,}99}} = \sqrt{\ln 10^2 n} \approx 2{,}14\sqrt{n}. \quad (31)$$

Пример 2. Пусть в качестве хэш-функции используется хэш-функция SHA-1 и n=2^{160}.

Тогда

$$k_{0,5} = 0.83\sqrt{2^{160}} = 0{,}83 \cdot 2^{80} = 10^{24};$$

$$k_{0,99} = \sqrt{2^{160}} = 2{,}14 \cdot 2^{80} = 2{,}6 \cdot 10^{24}.$$

Полученные в результате решения второй задачи результаты позволяют сделать следующие выводы.

Если на вход преобразования H (вычислителя значений хэш-функции) подаются k случайных значений M, причем они образуют множество X, а затем еще раз k случайных M, что образует множество Y, то для того чтобы значения x_i и y_j множеств X и Y совпали с вероятностью $P_з$, необходимо выполнить

$$k = \sqrt{n \ln\left(\frac{1}{1-P_з}\right)}$$

экспериментов.

Проведем оценки значений k и $P_з$ для используемого в Украине алгоритма симметричного блокового шифрования ГОСТ 28147-89. В четвертом режиме длина криптографической контрольной сумы (имитовставки) равняется 64 битам. Поэтому m=64 и n=2^m=2^{64}. При $P_з$=0,99, используя (19), получим

$$k = 1{,}41\sqrt{\ln 10^2 n} = 1{,}41\sqrt{\ln 10^2 \cdot 2^{64}} \approx 1{,}3 \cdot 10^{10}.$$

При $P_з$=0,1, используя (19), получаем

$$k = 1{,}41\sqrt{\ln\frac{1}{0{,}9} n} = 1{,}41\sqrt{\ln 1{,}1 \cdot 2^{64}} \approx 1{,}8 \cdot 10^{9}.$$

При $P_з$=0,5, используя (18), имеем

$$k = 1{,}41\sqrt{n} = 1{,}41\sqrt{2^{64}} \approx 1{,}41 \cdot 2^{32} \approx 6{,}06 \cdot 10^{9}.$$

С учетом требований к криптографическим алгоритмам блокового шифрования, изложенных в [3], можно сделать вывод, что режим 4 выработки имитовставки ГОСТ 28147-89 существенно подвержен возможности создания коллизий.

Заключение

Полученное в наиболее общем виде уравнение (15) позволяет точно решить задачу определения количества экспериментов k, которые необходимо выполнить для создания коллизии с вероятностью $P_з$ на множестве объема n. Достаточно хорошим приближением оценки k есть соотношение (19). Вместе с тем принимаемая в ряде источников без пояснения оценка $\sqrt{n}$ дает грубый результат. Наиболее точное значение можно получить из решения уравнения (15).

Выражение (29) позволяет оценить условия коллизий между двумя независимыми экспериментами в зависимости от размера n множества выходов и вероятности, с которой коллизия должна состоятся. Соотношения (30) и (31) дают приближенные оценки.

Полученные в результате решения задач результаты позволяют получить как зависимость k от $P_з$ и n, так и зависимость $P_з$ от k и n. Указанное может использоваться при синтезе и анализе алгоритмов различных криптографических преобразований, например, блочных криптографических хэш-функций.

Алгоритм ГОСТ 28147-89 в 4 режиме не обеспечивает защиту от коллизий.

Список литературы: 1. *В. Столлингс.* Криптография и защита сетей. Принципы и практика. 2-е изд-е.. К.: Изд. дом "Вильямс". 2001, 669 с. 2. *А.П. Алферов.* Основы криптографии. М.: Гелиос АРВ, 2001. 478 с. 3. *А.А. Молдовян.* Криптография. С.-Петербург. 2001. 218 с.

Харьковский национальный университет радиоэлектроники

Поступила в редколлегию 17.04.2002