## 5.3 Парадокс дня рождения

Предположим, количество выходных значений *хэш-функции* $H$ равно $n$. Каким должно быть число $k$, чтобы для конкретного значения $X$ и значений $Y_1, \ldots, Y_k$ вероятность того, что хотя бы для одного $Y_i$ выполнялось равенство

H (X) = H (Y) была бы больше 0,5.

Для одного $Y$ вероятность того, что $H(X) = H(Y)$, равна $1/n$.

Соответственно, вероятность того, что $H(X) \neq H(Y)$, равна $1 - 1/n$.

Если создать $k$ значений, то вероятность того, что ни для одного из них не будет совпадений, равна произведению вероятностей, соответствующих одному значению, т.е. $(1 - 1/n)^k$.

Следовательно, вероятность, по крайней мере, одного совпадения равна

$1 - (1 - 1/n)^k$

По формуле бинома Ньютона

$(1 - a)^k = 1 - ka + (k(k-1)/2!)a^2 - ... \approx 1 - ka$

$1 - (1 - k/n) = k/n = 0,5$

$k = n/2$

Таким образом, мы выяснили, что для m-битового *хэш-кода* достаточно выбрать $2^{m-1}$ сообщений, чтобы вероятность совпадения *хэш-кодов* была больше 0,5.

Теперь рассмотрим следующую задачу: обозначим P (n, k) вероятность того, что в множестве из k элементов, каждый из которых может принимать n значений, есть хотя бы два с одинаковыми значениями. Чему должно быть равно k, чтобы P (n, k) была бы больше 0,5?

Число различных способов выбора элементов таким образом, чтобы при этом не было дублей, равноn(n-1) … (n-k+1)n!/(n-k)!

Всего возможных способов выбора элементов равно $n^k$

Вероятность того, что дублей нет, равна $n!/(n-k)!n^k$

Вероятность того, что есть дубли, соответственно равна $1 - n!/(n-k)!n^k$

$P(n, k) = 1 - n! / ((n-k)! \times n^k) = 1 - (n \times (n-1) \times \ldots \times (n-k-1)) / n^k =$

$1 - [ (n-1)/n \times \ldots \times (n-k+1)/n] = 1 - [(1- 1/n) \times (1 - 2/n) \times \ldots \times (1 - (k-1)/n)]$

Известно, что $1 - x \leq e^{-x}$

$P(n, k) > 1 - [e^{-1/n} \times e^{-2}/n \times \ldots \times e^{-k}/n]$

$P(n, k) > 1 - e^{-k(k-1)/n}$

$1/2 = 1 - e^{-k(k-1)/n}$

$2 = e^{k(k-1)/n}$

$\ln 2 = k(k-1) / 2n$

$k(k-1) \approx k^2$

$k = (2n \times \ln 2)^{1/2} = 1,17\ n^{1/2} \approx n^{1/2}$

Если *хэш-код* имеет длину m бит, т.е. принимает $2^m$ значений, то

$k = \sqrt{2m} = 2^{m/2}$

Подобный результат называется "парадоксом дня рождения", потому что в соответствии с приведенными выше рассуждениями для того, чтобы вероятность совпадения дней рождения у двух человек была больше 0,5, в группе должно быть всего 23 человека. Этот результат кажется удивительным, возможно, потому, что для каждого отдельного человека в группе вероятность того, что с его днем рождения совпадет день рождения кого-то другого в группе, достаточно мала.

Вернемся к рассмотрению свойств *хэш-функций*. Предположим, что используется 64-битный *хэш-код*. Можно считать, что это вполне достаточная и, следовательно, безопасная длина для *хэш-кода*. Например, если зашифрованный *хэш-код* C передается с соответствующим незашифрованным сообщением M, то противнику необходимо будет найти M' такое, что

H (M') = H (M)

для того, чтобы подменить сообщение и обмануть получателя. В среднем противник должен перебрать $2^{63}$ сообщений для того, чтобы найти такое, у которого *хэш-код* равен перехваченному сообщению.

Тем не менее, возможны различного рода атаки, основанные на "парадоксе дня рождения". Возможна следующая стратегия:

1. Противник создает $2^{m/2}$ вариантов сообщения, каждое из которых имеет некоторый определенный смысл. Противник подготавливает такое же количество сообщений, каждое из которых является поддельным и предназначено для замены настоящего сообщения.
2. Два набора сообщений сравниваются в поисках пары сообщений, имеющих одинаковый *хэш-код*. Вероятность успеха в соответствии с "парадоксом дня рождения" больше, чем 0,5. Если соответствующая пара не найдена, то создаются дополнительные исходные и поддельные сообщения до тех пор, пока не будет найдена пара.
3. Атакующий предлагает отправителю исходный вариант сообщения для подписи. Эта подпись может быть затем присоединена к поддельному варианту для передачи получателю. Так как оба варианта имеют один и тот же *хэш-код*, будет создана одинаковая подпись. Противник будет уверен в успехе, даже не зная ключа шифрования.