А. Ю. Зубов (к. ф.-м. н., доцент)
ИКСИ

ОЦЕНКА СТОЙКОСТИ КОДА АУТЕНТИФИКАЦИИ С ДВУМЯ СОСТОЯНИЯМИ ИСТОЧНИКА ПРИ СЛУЧАЙНОМ И РАВНОВЕРОЯТНОМ ВЫБОРЕ КЛЮЧЕЙ

Получены выражения вероятностей успеха имитации и подмены через параметры матрицы кодирования для кодов аутентификации с двумя состояниями источника и равновероятным выбором ключей.

Код аутентификации (*А-код*) — это математическая модель системы аутентификации информации, основная цель которой — обеспечение защиты (хранимых, обрабатываемых или передаваемых) данных от активных атак. К активным атакам относятся действия злоумышленника по фальсификации или модификации данных. Обычно эти действия называют соответственно попытками имитации и подмены. Термин «код аутентификации» введен Г. Симмонсом в начале 70-х годов XX века. Разработке конструкций A-кодов и оценке их стойкости посвящено большое число работ. Наиболее полный список публикаций по этой тематике приведен в монографии [1].

Стойкость A-кодов обеспечивается использованием секретных ключей (как и при симметричном шифровании) и оценивается величиной вероятностей успеха активных атак. Чем меньше эта вероятность, тем более стоек A-код к соответствующей атаке. Используются различные определения вероятностей успеха активных атак. Одни предусматривают случайный выбор ключей в соответствии со специальным распределением вероятностей (называемым *оптимальной стратегией защиты*), которое, вообще говоря, может отличаться от равномерного (см. [2]), другие предусматривают лишь случайный и равновероятный выбор ключей. Для одних A-кодов вероятности успеха активной атаки, вычисленные в соответствии с разными определениями, могут незначительно различаться или даже совпадать, для других A-кодов соответствующие вероятности могут существенно различаться [3]. Каждое определение имеет свои преимущества и недостатки. Основным преимуществом определений, использующих равновероятный выбор ключей, является то, что значительно упрощается оценка вероятности успеха активной атаки. Однако такая оценка может быть завышенной по сравнению со значением вероятности, соответствующей оптимальной стратегии защиты. Вместе с тем, если A-код удовлетворяет условию стойкости при равновероятном выборе ключей, то он будет удовлетворять условию стойкости и при использовании оптимальной стратегии защиты. Это оправдывает использование равновероятного выбора ключей A-кода во многих практических приложениях.

В данной статье выражаются значения вероятностей успеха атак имитации и подмены для A-кодов с минимально возможным числом состояний источника и равновероятным выбором ключей через параметры матрицы кодирования.

Напомним основные определения. Пусть S, M, E — множества *состояний источника*, *сообщений* и *правил кодирования* соответственно. Каждое правило кодирования представляет собой инъективное отображение $e{:}S\to M$. Оно определяет обратное отображение $e^{-1}{:}M\to S\cup\{0\}$ формулой:

$$e^{-1}(m)=\begin{cases} s\in S, & \text{если } m\in e(S), \text{ причём } m=e(s),\\ 0, & \text{если } m\notin e(S), \end{cases}$$

где $e(S)=\{e(s)\colon s\in S\}$. При этом полагают, что $0\notin S$. Из условия инъективности правила кодирования следует, что для любых $s\in S$ и $e\in E$ выполняется равенство:

$$e^{-1}(e(s))=s, \tag{1}$$

означающее, что сообщение и правило кодирования однозначно восстанавливают состояние источника. Полагают, чтобы выполнялось равенство:

$$M=\bigcup_{e\in E} e(S), \tag{2}$$

означающее, что каждое сообщение является результатом применения некоторого правила кодирования к некоторому состоянию источника.

Код аутентификации — это тройка конечных множеств (S, E, M), где $|S|\geq 2$, $|E|\geq 3$, $|M|\geq 3$, удовлетворяющих условиям (1) и (2). Для удобства выбора правил кодирования множество E индексируется *ключами*. Для аутентификации передаваемых состояний источника передатчик и приемник выбирают (в секрете от оппонента) общее правило кодирования e. Передатчик вычисляет $m=e(s)$ и направляет сообщение m приемнику. Критерием аутентичности полученного сообщения является условие $e^{-1}(m)\neq 0$.

Состояния источника появляются случайно в соответствии с заданным распределением вероятностей $P(S)=(\rho_S(s),\ s\in S\}$. Пусть $\tilde{S}$ — случайная величина, определенная для любого $s\in S$ равенством $P\{\tilde{S}=s\}=\rho_S(s)$. Передатчику и приемнику целесообразно выбирать правила кодирования случайно. При этом они принимают решение о введении на множестве E некоторого распределения вероятностей $P(E)=(\rho_E(e), e\in E)$ (стратегия защиты). Пусть $\tilde{E}$ — соответствующая случайная величина. Полагается, что случайные величины $\tilde{S}$ и $\tilde{E}$ независимы. Они индуцируют случайную величину $\tilde{M}$ с множеством исходов M. Вероятность $\rho_M(m)$ вычисляется по формуле:

$$p_M(m)=\sum_{e\in E(m)} p_E(e)\cdot p_S(e^{-1}(m)), \qquad (3)$$

где $E(m)=\{e\in E: e^{-1}(m)\neq 0\}$.

A-код можно задать $|E|\times|M|$ — матрицей, называемой *матрицей кодирования*. Строки матрицы занумерованы правилами кодирования $e\in E$, столбцы — сообщениями $m\in M$; на пересечении строки с номером e и столбца с номером m расположен элемент $e^{-1}(m)$. Помимо матрицы кодирования вводится *матрица инцидентности* A-кода — X_A, имеющая те же размеры. Ее элементы $x(e,m)$ определяются формулой:

$$x(e,m)=\begin{cases}1, \text{ если } e^{-1}(m)\neq 0,\\ 0, \text{ если } e^{-1}(m)=0.\end{cases}$$

Естественно потребовать ([1]), чтобы матрица X_A не содержала нулевых строк и столбцов, содержала хотя бы один ноль и не менее двух единиц в каждой строке и каждом столбце и не содержала двух и более одинаковых столбцов.

Будем оценивать стойкость A-кода к атаке имитации вероятностью ρ_0, определяемой следующим образом. Пусть $P(E)$ — равномерное распределение и $\rho_и(m)$ — вероятность события "$m\in e(S)$" при случайном выборе $e\in E$. Это вероятность того, что сообщение $m\in M$ будет принято как аутентичное. Она выражается формулой:

$$\rho_и=(m)=\Sigma_{e\in E(m)}\rho_E(e).$$

Тогда ρ_0 определим как

$$p_0=\max_{m\in M} p_и(m)=\max_{m\in M}|E(m)|/|E|. \qquad (4)$$

Стойкость A-кода к атаке подмены будем оценивать вероятностью ρ_1, определяемой следующим образом. Пусть $P(E)$ — равномерное распределение, и $m,n\in M$ — различные сообщения. Пусть $\rho_п(n/m)$ — вероятность того, что для случайно выбранной пары (e,s) оппонент добьется успеха при подмене наблюдаемого сообщения $m\in M$ сообщением $n\neq m$. Эта вероятность вычисляется по формуле:

$$\rho_п(n/m)=1/\rho_M(m)\cdot\sum_{e\in E(m,n)}\rho_E(e)\rho_S(e^{-1}(m)),$$

где $E(m,n)=E(m)\cap E(n)$. Пусть

$$\rho_п=(m)=\max_{n\neq m}\rho_п(n/m).$$

Тогда

$$\rho_1=\sum_{m\in M}\rho_M(m)\rho_п(m)=\sum_{m\in M}\max_{n\neq m}\sum_{e\in E(m,n)}\rho_E(e)\rho_S(e^{-1}(m))=1/|E|\cdot\sum_{m\in M}\max_{n\neq m}\sum_{e\in E(m,n)}\rho_S(e^{-1}(m)). \qquad (5)$$

В частном случае, когда распределение $P(S)$ равномерно, формула (5) принимает вид:

$$p_1=\frac{1}{|E|\cdot|S|}\sum_{m\in M}\max_{n\neq m}|E(m,n)|. \qquad (6)$$

Определение вероятности успеха подмены по формулам (5) и (6) использует подход «в среднем». Помимо этого, используется и подход «в худшем случае», когда вероятность успеха подмены определяется как

$$p_1' = \max_{m\in M}\max_{n\neq m} p_{\text{п}}(n/m) = \max_{m\in M}\max_{n\neq m}\sum_{e\in E(m,n)} p_S(e^{-1}(m)) : \sum_{e\in E(m)} p_S(e^{-1}(m)). \qquad (7)$$

В частном случае, когда и распределение $P(S)$ равномерно:

$$p_1' = \max_{m\in M}\max_{n\neq m} |E(m,n)| / |E(m)|. \qquad (8)$$

Ясно, что выполняется неравенство $\rho_1' \geq \rho_1$

Известны (см. [1]) достижимые нижние оценки этих вероятностей:

$$\rho_0 \geq \frac{|S|}{|M|},\ \rho_1 \geq \frac{|S|-1}{|M|-1},\ \max\{\rho_0, \rho_1\} \geq \frac{1}{\sqrt{|E|}}\ . \qquad (9)$$

Оценки (9) называются комбинаторными. Имеются также энтропийные оценки вида:

$$\log_2\rho_0 \geq H(\tilde{E}/\tilde{M}) - H(\tilde{E}) \text{ и } \log_2\rho_1 \geq -H(\tilde{E}/\tilde{M}), \qquad (10)$$

где $H(\tilde{E})$, $H(\tilde{E}/\tilde{M})$ — энтропия и условная энтропия случайной величины $\tilde{E}$.

Получим новые комбинаторные оценки вероятностей ρ_0, ρ_1, ρ_1' для A-кодов с двумя состояниями источника. Такие A-коды могут быть использованы, например, при передаче сообщений о результатах случайного бросания монеты.

Пусть $S=\{H,T\}$, $P(S)=(\rho_S(H)=\rho, \rho_S(T)=1-\rho)$, причем $\rho\geq 0{,}5$. Обозначим $|M|=v$, $|E|=b$. Пусть A — матрица кодирования A-кода. Она имеет размеры $b\times v$. В каждой ее строке имеется ровно 2 ненулевых элемента — H и T. Пусть $M_H(M_T)$ — подмножество сообщений, которым соответствуют столбцы матрицы A, состоящие из элементов $\{0,H\}$ ($\{0,T\}$). Обозначим $|M_T|=t$.

Заметим, что для A-кода с двумя состояниями источника для любых сообщений $m,n\in M$, $n\neq m$, выполняются неравенства $2\leq|E(m)|$, $|E(m,n)|\leq 2$, причем $|E(m,n)|=2$ в том и только в том случае, если $e_1(H)=e_2(T)=m$, $e_1(T)=e_2(H)=n$. Пусть

$$M_{2(H,T)}=\{m\in M: \exists n\in M(|E(m,n)|=2)\},$$

и $|M_{2(H,T)}|=\lambda$. Очевидно, что $\lambda\leq b/2$.

При подсчете числа ненулевых элементов матрицы A по строкам и по столбцам получаем формулу:

$$\sum_{m\in M}|E(m)|=2b$$

из которой следует, что

$$\max_{m\in M}|E(m)|\geq 2b/v,$$

и $\rho_0\geq 2/v$. Это совпадает с неравенством, указанным в (9). Из того, что $2\leq|E(m)|$, попутно получаем неравенство $v\leq b$, означающее, что для A-кода с двумя состояниями источника число правил кодирования должно быть не меньше числа сообщений. Отметим также, что равенство $\rho_0=2/v$ выполняется в том и только в том случае, если $|E(m)|=2b/v$ для любого сообщения $m\in M$. В частности, $2b$ делится на v.

Получим теперь формулу для ρ_1.

Рассмотрим слагаемое из суммы, указанной в (5). Оно представляется в виде:

$$\rho_M(m)\rho_{\text{п}}(m)=(1/b)\cdot\max_{n\neq m}\sum_{e\in E(m,n)}\rho_S(e^{-1}(m)).$$

В силу сказанного выше:

$$\sum_{e\in E(m,n)}\rho_S(e^{-1}(m)) = \begin{cases} 0, \text{ если } |E(m,n)|=0, \\ \rho, \text{ если } |E(m,n)|=1 \text{ и } m\in M_H, \\ 1-\rho, \text{ если } |E(m,n)|=1 \text{ и } m\in M_T, \\ 1, \text{ если } |E(m,n)|=2, \end{cases}$$

и

$$\rho_M(m)\rho_{\text{п}}(m) = \begin{cases} 1/b, & \text{если } m\in M_{2(H,T)}, \\ \rho/b, & \text{если } m\notin M_T\cup M_{2(H,T)}, \\ (1-\rho)/b, & \text{если } m\in M_T. \end{cases}$$

Отсюда получаем:

Утверждение 1. *Для A-кода с двумя состояниями источника вероятность* ρ_1 *успеха подмены выражается формулой:*

$$\rho_1=\begin{cases} v/b, & \text{если } M_{2(H,T)}=M, \\ 1/b\ (2\lambda + t(1-\rho)+(v-2\lambda-t)\rho), & \text{если } M_{2(\mathrm{H,T})}\neq M. \end{cases} \quad (11)$$

Следствие 1. *Если для A-кода с двумя состояниями источника не более одного правила кодирования допустимо для любой пары сообщений* ($\lambda=0$), *то*

$$\rho_1=(1/b)(v\rho-t(2\rho-1)). \quad (12)$$

Если дополнительно матрица A *не содержит столбцов, состоящих лишь из элементов* $\{0,T\}$, ($\lambda=t=0$), *то* $\rho_1=v\rho/b$.

Заметим, что во втором случае следствия величина $1/\rho_1$ представляет собой среднее число ненулевых элементов в столбце матрицы A.

Следствие 2. *Для A-кода с двумя равновероятными состояниями источника* ($\rho=0{,}5$):

$$\rho_1=v/2b.$$

Полученные утверждения позволяют по внешнему виду матрицы кодирования легко вычислять вероятности ρ_0 и ρ_1. Рассмотрим примеры.

Пример 1. Пусть матрица A имеет вид:

$$\begin{bmatrix} H\ T\ 0\ 0\ .\ 0\ 0 \\ H\ 0\ T\ 0\ .\ 0\ 0 \\ H\ 0\ 0\ T\ .\ 0\ 0 \\ .\ .\ .\ .\ .\ .\ . \\ H\ 0\ 0\ 0\ .\ T\ 0 \\ 0\ T\ 0\ 0\ .\ 0\ H \\ 0\ 0\ T\ 0\ .\ 0\ H \\ 0\ 0\ 0\ T\ .\ 0\ H \\ .\ .\ .\ .\ .\ .\ . \\ 0\ 0\ 0\ 0\ .\ T\ H \end{bmatrix}.$$

Для такого A-кода выполняются соотношения $v=(b/2)+2=$, $t=v-2=b/2$, $\lambda=0$, из которых получаем значение $\rho_1=0{,}5-\rho(1-v/b)$.

В примере 1 величина t принимает (при фиксированных параметрах b и v) максимально возможное значение, поэтому вероятность ρ_1, вычисленная по формуле (12), принимает минимально возможное значение. С ростом числа ключей эта вероятность стремится к 0. Вместе с тем легко увидеть, что $\rho_0=0{,}5$, поэтому такой A-код не является стойким. Приведем пример стойкого A-кода.

Пример 2. Пусть v – произвольное натуральное число. Рассмотрим A-код с двумя состояниями источника, матрица кодирования которого содержит максимально возможное число строк, равное числу размещений $A_v^2=v(v-1)=b$. Ясно, что каждый столбец такой матрицы содержит одинаковое число ненулевых элементов – $|E(m)|=2b/v=2(v-1)$, откуда следует, что ρ_0 принимает минимально возможное значение, равное $2/v$. Для данного A-кода выполняется равенство $M_{2(H,T)}=M$, поэтому согласно (11) и ρ_1 принимает минимально возможное значение $\rho_1=v/b=1/(v-1)$. Если считать стойким A-код, для которого выполняется неравенство $\max\{\rho_0,\rho_1\}\le\varepsilon$, где ε– выбранный порог стойкости, то для предложенного A-кода неравенство выполняется при $v\ge 2/\varepsilon$. Например, для $\varepsilon=2^{-20}$ стойкость обеспечивается при $b=2^{21}(2^{21}-1)\approx 2^{42}$. Для выбора секретного правила кодирования требуется 42 бита ключа.

Получим формулу для вычисления вероятности ρ_1'. Обозначим через h_m число вхождений состояния источника H в столбец матрицы A с номером m. Напомним, что A-код называется *A-кодом без секретности* (кратко – *$\bar{A}$-кодом*), если $M=M_H\cup M_T$. В противном случае A-код называется *A-кодом с секретностью*. Наибольшее распространение в практических приложениях получили *$\bar{A}$-коды*.

Заметим, что

$$\max_{n\neq m}\frac{\sum\limits_{e\in E(m,n)} p_S(e^{-1}(m))}{\sum\limits_{e\in E(m)} p_S(e^{-1}(m))}=\begin{cases}\dfrac{1}{|E(m)|}, & \text{если } m\in M_H\cup M_T,\\ \dfrac{p}{h_m p+(|E(m)|-h_m)(1-p)}, & \text{если } m\notin M_{2(H,T)}\cup M_H\cup M_T,\\ \dfrac{1}{h_m p+(|E(m)|-h_m)(1-p)}, & \text{если } m\in M_{2(H,T)}.\end{cases} \qquad (13)$$

Введем следующие обозначения:

$$F_1(m)=\frac{1}{|E(m)|},\quad F_2(m)=\frac{p}{h_m p+(|E(m)|-h_m)(1-p)},\quad F_3(m)=\frac{1}{h_m p+(|E(m)|-h_m)(1-p)}.$$

Из (7) и (13) следует:

Утверждение 2. *Для A-кода с двумя состояниями источника вероятность ρ'_1 успеха подмены выражается формулой:*

$$\rho'_1=\max\Big\{\max_{m\in M_H\cup M_T} F_1(m),\ \max_{m\notin M_{2(H,T)}\cup M_H\cup M_T} F_2(m),\ \max_{m\in M_{2(H,T)}} F_3(m)\Big\}. \qquad (14)$$

Следствие 1. *Для $\bar{A}$-кода с двумя состояниями источника*

$$\rho'_1=\max_{m\in M}1/|E(m)|.$$

Следствие 2. *Для A-кода с двумя равновероятными состояниями источника* ($\rho=0{,}5$)

$$\rho'_1=\max\Big\{\max_{m\notin M_{2(H,T)}} F_1(m),\ \max_{m\in M_{2(H,T)}} 2F_1(m)\Big\}.$$

Следствие 3. *Если для $\bar{A}$-кода с двумя состояниями источника $\rho_0=2/v$, то $\rho'_1=v/2b$.*

Заметим, что A-код из примера 1 является $\bar{A}$-кодом. Согласно следствию 1 для него $\rho'_1=0{,}5$, в то время как $\rho_1=0{,}5-\rho(1-v/b)$. Как видим, разница между значениями ρ_1 и ρ'_1 может приближаться к 0,5. A-код из примера 2 не является $\bar{A}$-кодом. Для него выполняются соотношения $|E(m)|=2(v-1)$, $h_m=v-1$, для любого $m\in M$, и $M=M_{2(H,T)}$, откуда следует, что $\rho'_1=1/(v-1)$. Для этого A-кода ρ_1 и ρ'_1 совпадают.

СПИСОК ЛИТЕРАТУРЫ

1. *Зубов А. Ю.* Математика кодов аутентификации. М.: Гелиос АРВ, 2007.
2. *Зубов А. Ю.* К теоретико-игровому подходу исследования кодов аутентификации // Дискретная математика. 2008.
3. *Зубов А. Ю.* О выборе оптимальной стратегии защиты для кода аутентификации с двумя состояниями источника // Дискретная математика. 2008.