УДК 681.3.06

Г. З. ХАЛИМОВ, канд. техн. наук

БЕЗУСЛОВНАЯ АУТЕНТИФИКАЦИЯ С ИСПОЛЬЗОВАНИЕМ СЛАБО СМЕЩЕННЫХ МАССИВОВ

Безусловная аутентификация, предложенная Г.Ж. Симмонсом [1, 2], основывается на использовании семейства почти строго универсальных хэш-функций. Аутентификация Симмонса снимает ограничение на большую размерность ключевых данных в конструкции кодов аутентификации Картера-Вергмана [3, 4]. Семейство почти строго универсальных хэш-функций можно определить в рамках теории мало смещённых, почти независимых массивов (weakly biased arrays, almost independent arrays). Определение смещённых массивов введено в работах [5, 6] для массивов дискретных значений большой размерности с распределением незначительно отличающимся от равномерного. Хеллесет и Джохансон впервые установили взаимосвязь теории кодирования и смещённых массивов [7]. Применение в схемах аутентификации незначительно смещённых, почти независимых массивов с использованием алгеброгеометрических кодов, как будет показано в статье, позволяет получить при заданном объёме ключа и данных источника сообщений вероятность коллизии не больше, чем в случае использования строго универсальных хэш-функций.

Задачей данной статьи является изложение общетеоретических вопросов построения аутентификации с применением почти строго универсального хэширования на основе слабо смещенных, почти независимых массивов в конструкции с алгебраическими кодами. С этой целью в разделе 1 приводятся определения смещённых и зависимых массивов. В разделе 2 рассматривается применение теории кодирования для построения мало смещённых массивов в схемах аутентификации.

1 Определение смещённых и зависимых массивов

Понятия ограниченной зависимости, слабого смещения используются в различных приложениях криптографии и теории сложности: при аутентификации, универсальном хэшировании, оценке устойчивости против корреляционных атак, псевдорандомизации, тестировании комбинаторных схем и др. Определения ε -смещенных массивов, t -связных и ε – зависимых массивов приведём в изложении Биербрауэра [8].

Определение 1. $(n,k)_p$-массив, содержащий n строк, k столбцов и записи из набора p элементов.

Определение 2. Пусть p - простое число, $u=(u_1,u_2,...,u_n)\in F_p^n$. Для $\forall i\in F_p$, $v_i(u)$ есть частота появления элемента i в последовательности u $v_i(u)=\frac{n}{p}+\delta_i(u)$, где $\delta_i(u)$ – есть отклонение частоты $v_i(u)$ от среднего значения и $\sum_{i\in F_p}\delta_i(u)=0$. Пусть ξ комплексный корень p - степени из единицы, тогда смещение вектора u определяется как

$$bias(u)=\frac{1}{n}\left|\sum_{i\in F_p}\delta_i(u)\xi^i\right|=\frac{1}{n}\left|\sum_{i\in F_p}v_i(u)\xi^i\right|.$$

Докажем следующее полезное утверждение.

Утверждение 1. Для произвольного вектора u $0\le bias(u)\le 1$ и $bias(u)=1$ только тогда, когда $u=const$.

Действительно, по определению нормы $bias(u)$ не может быть меньше нуля. Так как $\sum_{i\in F_p} \delta_i(u) = 0$, имеем

$$\frac{1}{n}\left|\sum_{i\in F_p} v_i(u)\xi^i\right| = \frac{1}{n}\left|\sum_{i\in F_p}\left(\frac{n}{p}+\delta_i(u)\right)\xi^i\right| = \frac{1}{n}\left|\sum_{i\in F_p}\delta_i(u)\xi^i\right|.$$

Для произвольного вектора в силу того, что $|\delta_i(u)| \le \frac{n}{p}$, получим

$$bias(u) = \frac{1}{n}\left|\sum_{i\in F_p}\delta_i(u)\xi^i\right| \le \frac{1}{n}\sum_{i\in F_p}\left|v_i(u)\xi^i\right| \le \frac{1}{n}\sum_{i\in F_p}|\delta_i(u)|\left|\xi^i\right| \le 1.$$

Если $u = const$ тогда $bias(u) = \frac{1}{n}\left|\sum_{i\in F_p}\delta_i(u)\xi^i\right| = \frac{1}{n}\left|n\xi^i\right| = 1.$

Определение 3. Пусть $0 \le \varepsilon \le 1$. Массив $(n,k)_p$ является ε -смещённым (-biased), если любая нетривиальная линейная комбинация столбцов имеет смещение $bias \le \varepsilon$.

Смещение массива является свойством F_p – линейного кода, построенного с помощью столбцов порождающей матрицы. Следующие определения являются обобщением понятия ортогональных массивов силы t. В отечественной литературе с ортогональными массивами отождествляются t -схемы, ортогональные таблицы [9], t-универсальные семейства хэш-функций [10].

Определение 4. Пусть $0 \le \varepsilon \le 1$. Массив $(n,k)_p$ является t – связным (t – wise), ε - смещённым, если любая нетривиальная линейная комбинация не более чем t столбцов имеет смещение $bias \le \varepsilon$.

Определение 5. Пусть $0 \le \varepsilon \le 1$. Массив $(n,k)_p$ является t – связным, ε - зависимым (ε - dependent), если для любого набора U из $s \le t$ столбцов и каждого вектора $a \in F_p^s$ частота $v_U(a)$ появления в столбцах значения a удовлетворяет условию

$$\left|\frac{v_U(a)}{n} - \frac{1}{p^s}\right| \le \varepsilon.$$

Пример 1. Рассмотрим двоичный (n,k) код, ненулевые слова которого имеют вес, удовлетворяющий условию

$$\frac{1-\varepsilon}{2} \le \frac{\omega_i}{n} \le \frac{1+\varepsilon}{2}, \qquad \varepsilon < 1.$$

Транспонированная порождающая матрица кода определяет массив $(n,k)_2$ со смещением

$$bias(u) = \frac{1}{n}\left|\sum_{i\in F_2}\delta_i(u)\xi^i\right| = \frac{1}{n}\left|\delta_0\xi^0 + \delta_1\xi^1\right| = \frac{1}{n}\left|-\frac{\varepsilon n}{2} - \frac{\varepsilon n}{2}\right| = \varepsilon.$$

Пример 2. Рассмотрим t – связный, независимый (0- зависимый) массив $(n,k)_p$. По определению 5 имеем $\frac{v_U(a)}{n} = \frac{1}{p^t}$. В этом случае $(n,k)_p$ является ортогональным массивом силы t и образует t – строго универсальное семейство хэш-функций.

Пример 3. Рассмотрим семейство ε – почти строго универсальных $_2$ (ASU $_2$) хэш – функций $h \in H$. По определению (см. [10]) имеем:

1. Для любых значений $x, y \in A$, $x \neq y$ число хэш-функций h таких, что $h(x) = h(y)$ не больше чем $\varepsilon|H|$;
2. Для произвольных $x \in A, y \in B$, число функций таких, что $h(x) = y$, строго равно $|H|/|B|$.

Как следует из второго условия, массив хэш-значений определяет $(n,k)_p$ массив со смешением равным нулю.

Применение слабо смещённых и почти независимых массивов для целей аутентификации определяется тем, что можно построить большие наборы случайных переменных, которые являются почти статистически независимыми. Ниже приводятся конструкции таких кодов аутентификации.

2 Коды аутентификации со слабо смещёнными массивами

Методы построения массивов со смещением достаточно полно представлены в [5, 6, 12]. Хеллесет и Джохансон впервые установили взаимосвязь теории кодирования и смещённых массивов[6]. Наиболее полно эта взаимосвязь была исследована Биербрауэром [8]. Рассмотрим основные кодово-теоретические результаты.

Теорема 1. Пусть (n,k) двоичный код, минимальное расстояние кода d и максимальное D удовлетворяет условию

$$Min\left\{\frac{d}{n}, \frac{n-D}{n}\right\} \geq \frac{1-\varepsilon}{2},$$

где $0 \leq \varepsilon \leq 1$. Тогда транспонированная порождающая матрица кода является ε – смещённым $(n,k)_2$ массивом.

Очевидно следующее утверждение.

Утверждение 2. Пусть $(n,k+1)$ двоичный код C' образован (n,k,d) кодом C и единичным кодовым словом 1, тогда минимальное кодовое расстояние точно равно $Min\{d, n-D\}$.

Отсюда сразу следует практическая конструкция.

Теорема 2. Пусть $(n,k+1,d)$ двоичный код C', содержащий единицу 1 и $\frac{d}{n} \geq \frac{1-\varepsilon}{2}$, $0 \leq \varepsilon \leq 1$, тогда транспонированная порождающая матрица кода является ε – смещённым $(n,k)_2$ массивом.

Построение асимптотически хороших ε- смещённых $(n,k)_2$ массивов (в смысле достижения как можно большего отношения $\frac{k}{n}$ при фиксированном значении смещения ε) эквивалентно решению задачи нахождения асимптотически хороших кодов, содержащих единицу 1.

Для случая p -ичных кодов основной результат получен в [8].

Теорема 3. Пусть C' $(n,k+1,d)_p$ линейный код, содержащий подкод C и m слов веса n. Транспонированная порождающая матрица кода C является $(n,k)_p$- массивом со смещением $\varepsilon \leq p-1-p\frac{d}{n}$.

Пример 4. Пусть $(p,k,p-k+1)_p$ код Рида-Соломона. По предыдущей теореме получим $(n,k-1)_p$ со смещением $\varepsilon \le p-1-p\dfrac{p-k+1}{p} = p-1-p+k-1 = k-2$. Для РС кода размерности $k=2$ имеем 0 - смещенный $(n,1)_p$ массив.

В работе [5] предложено усиление для кодовых конструкций в ε - смещенных $(n,k)_p$ массивах.

Теорема 4. Пусть p - простое число и C – (n,k,d) код. Тогда существует массив $(pn,k)_p$, который имеет смещение $\varepsilon \le 1-\dfrac{d}{n}$.

Пример 5. Рассмотрим $(p,k)_p$ код РС. Применяя теорему 4, получим $\dfrac{k-1}{p}$ – смещенный $\left(p^2,k\right)_p$ массив.

Дальнейшее обобщение теоремы 4 получено в [8].

Теорема 5. Пусть C линейный $(n,k,d)_{q=p^m}$ код и B внутренний $(n_0,m)_p$ массив со смещением ε_0. Тогда существует $(nn_0,km)_p$ массив со смещением $\varepsilon = 1-\delta+\delta\varepsilon_0 \le 1-\delta+\varepsilon_0$, где $\delta = \dfrac{d}{n}$.

Для построения внутренних массивов $(n_0,m)_p$ можно использовать несколько относительно простых конструкций [5].Несколько лучшие результаты имеет метод сумм экспонент Вейля- Карлитца- Ушиямы (ВКУ) [13]. Метод ВКУ состоит в построении массива A с записями вида $Tr\left(a_j\alpha^i\right)$, где a_j - базис поля $F_{p^f} \mid F_p$, $i \le n$ и i не кратно p, $Tr: F_{p^f} \to F_p$ – след элемента $a_j\alpha^i$. Строки массива A индексируются элементами $\alpha \in F_{p^f}$, а столбцы-функциями a_jX^i. Результирующий массив имеет параметры $\left(p^f, f*(n-n/p)\right)_p$ и смещение $bias \le (n-1)p^{-f/2}$, где $(n-n/p)$ определяет возможное число экспонент a_jX^i при $i \le n$ и i не кратно p.

Пример 6. Построим массив ВКУ $\left(p^f, f*(n-n/p)\right)_p$ со смещением $bias \le (n-1)p^{-f/2}$ при $p=2, f=4, n=1$. Базисные элементы поля имеют вид $a_j: 1,\alpha,\alpha^2,\alpha^3$. Так как $n=1$, следует взять только одну экспоненту $\varphi: X$. Строки массива индексируются элементами $\alpha \in F_{2^4}$, столбцы – функциями: $X, \alpha X, \alpha^2 X, \alpha^3 X$, а записи – $Tr(\beta) = \beta+\beta^2+\beta^4+\beta^8$. Получим $\left(2^4,4\right)$ массив со смещением $bias = (1-1)2^{-2} = 0$.

Пусть $p=3, f=2, n=2$. Тогда $a_j: 1,\alpha; \varphi: X, X^2; Tr(\beta) = \beta+\beta^3$. Строки массива индексируются элементами $\alpha \in F_{3^2}$ (порождающий многочлен поля z^2+z+2), столбцы – функциями : $X, \alpha X, \alpha X^2, X^2 = \alpha^4 X+1 \pmod{X^2+X+2}$. Массив $\left(3^2,4\right)_3$ имеет вид (см. табл.).

Таблица

α^i	X	αX	αX^2	$\alpha^4 X$
0	0	0	0	0
α^0	0	α^4	α^4	0
α^1	α^4	0	α^4	0
α^2	0	α^4	α^0	α^4
α^3	α^4	0	α^0	α^4
α^4	0	α^4	α^4	0
α^5	α^0	0	α^4	α^4
α^6	0	α^4	α^0	α^4
α^7	α^0	0	α^0	α^0

Зададим произвольную линейную комбинацию столбцов $Y = \sum_{j=1}^{4} \gamma^j Y_j$, $\gamma_j \in F_3, j = 1,4$, например, $Y = Y_1 + \alpha^4 Y_2 + \alpha^4 Y_4$. Получим результирующий вектор $Y_p = (0, \alpha^0, \alpha^4, \alpha^4, 0, \alpha^4, \alpha^4, 0, 0)$. Значения частот элементов $0, \alpha^0, \alpha^4$ равны: $\nu_0 = 4, \delta_0 = +1; \nu_{\alpha^0} = 1, \delta_{\alpha^0} = -2; \nu_{\alpha^4} = 4, \delta_{\alpha^4} = +1$, а смещение

$$bias(\nu_Y) = \frac{1}{9}\left|1 * e^{j\frac{2\pi}{3}*0} + (-2)e^{j\frac{2\pi}{3}*1} + 1 * e^{j\frac{2\pi}{3}*2}\right| = \frac{1}{9}\left|1 + 1 - \sqrt{3}j - \frac{1}{2} - \frac{\sqrt{3}}{2}j\right| = \frac{1}{3}.$$

Для всех нетривиальных линейных комбинаций столбцов значение $bias \le \frac{1}{3}$ удовлетворяет соотношению $bias \le p^{-1}$, при $n = 2, f = 2$.

Обобщая полученные результаты, отметим, что существует массив ВКУ с параметрами $(p^2, 4)_p$ и смещением $bias = \frac{1}{p}$, где p - простое.

Пример 7. Зададим в поле $F_{q=p^4}$ РС код с параметрами $(p^4, p^3, p^4 - p^3 + 1)$ и внутренний массив $(p^2, 4)_p$ (см. предыдущий пример). Тогда по теореме 5 получим массив $(p^6, 4p^3)_p$ со смещением $bias = 1 - \frac{p^4 - p^3 + 1}{p^4} + \frac{p^4 - p^3 + 1}{p^4} * \frac{1}{p} \le \frac{2}{p}$.

Метод ВКУ при тех же параметрах массива гарантирует смещение $bias = \left(\frac{4}{6}p^3 - 1\right)p^{-\frac{6}{2}} < \frac{4}{6}$ и это уступает каскадной схеме.

Рассмотрим t – связные массивы. Основная кодово-теоретическая конструкция описывается в [6].

Теорема 6. Пусть $(n,k)_p$ массив B со смещением ε и $(N, N-k, t+1)$ -линейный код. Тогда существует $(n,N)_p$ – массив, который является t – связным и ε – смещенным.

Результирующий массив можно построить путем умножения матрицы B и проверочной матрицы кода H. Важное соотношение между смещением и зависимостью установлено в [12].

Теорема 7. Если массив является t – связным и ε – смещенным, он является также и t – связным и ε' – зависимым, причём, $\varepsilon' < \varepsilon$.

Фундаментальное значение этой теоремы заключается в том, что она определяет возможность применения слабо смещённых массивов в схемах аутентификации.

Приведём основные определения для универсальных кодов аутентификации в терминологии слабо смещенных и почти независимых массивов.

Определение 6. $(n,k)_q$ – массив является ε – почти строго универсальным $_2$ (ASU_2), если его столбцы имеют смещение 0 и для различных столбцов C, C' и любых записей e, e', при равновероятном выборе i строки, условная вероятность $\Pr(c_i = e \mid c_i' = e') \le \varepsilon$.

Выбор строки массива определяется значением ключа, столбец – состоянием источника данных и значение записи является кодом аутентификации.

Определение 7. $(N,m)_p$ – массив является (δ,t) – почти строго универсальным $_t$ $(\delta - ASU_t)$, если для любого набора $U = U_0 \bigcup \{u\}$ из t столбцов и каждой записи $a' \in F_p{}^{t-1}, x \in F_p$ отношение частот $\nu_{U_0}(a')$ и $\nu_U(a', x)$ удовлетворяет условию

$$\left|\nu_U(a', x)/\nu_{U_0}(a')\right| \le \delta .$$

Связь между почти независимыми массивами и $\delta - ASU_t$ кодами достаточно очевидна и была установлена в [12].

Теорема 8. $\delta - ASU_t$ является t – связным ε – зависимым массивом, где $\delta = \left(p^{-t} + \varepsilon\right) / \left(p^{-(t-1)} - \varepsilon\right)$.

Пример 8. Построим массив аутентификаторов ASU_2. Зададим C – $(n,k)_p$ линейный код с ε_0 - смещёнными кодовыми словами. Строки массива ASU_2 индексируем набором $(i, \alpha_1, \alpha_2, ..., \alpha_t)$, где i - номер элемента C кода, $\alpha_r \in F_p$. Зададим линейные отображения $M_r : C \to C, r = 1,2,...,t$ так, что любая нетривиальная F_p - линейная комбинация из отображений M_r является несингулярной. Столбцы массива ASU_2 являются словами кода. Запись в массиве на пересечении строки $(i, \alpha_1, \alpha_2, ..., \alpha_t)$ и столбца f определяется выражением $(M_1(f)(i) + \alpha_1, M_2(f)(i) + \alpha_2, ..., M_t(f)(i) + \alpha_t)$. Смещение каждого столбца равно нулю, так как $(\alpha_1, \alpha_2, ..., \alpha_t)$ пробегают p^t значений. По определению 6 имеем $\delta - ASU_t$ $(np^t, p^k)_{p'}$ массив. Используя результат теоремы 7 и определение 5, для t - связных и ε - зависимых массивов получим границу $\varepsilon \le p^{-t} + \varepsilon_0$.

Пример 9. Рассмотрим $\left(p^3, p^2/2, p^3 - p^2\right)_{p^2}$ код Эрмита (см. [11]). Построим методом ВКУ несмещённый $\left(p^2, 2\right)_p$ массив. Каскадная конструкция кода Эрмита и несмещённого массива по теореме 5 дает $\left(p^5, p^2\right)_p$ - массив с $\varepsilon = \frac{1}{p}$. Используя результаты примера 8, получим дальнейшее улучшение $\rightarrow$ $\frac{2}{p}$ – $\left(p^6, p^{p^2}\right)$ массив ASU_2, что лучше, чем дает использование кодов РС (см. пример 7).

Применение алгеброгеометрических кодов высокого порядка в каскадных конструкциях со слабо смещёнными массивами дает лучшие результаты, что согласуется с выводами, сделанными в работе [11].

Существующее широкое многообразие кодово-теоретических схем построения смещённых и связных массивов позволяет исследовать коды аутентификации с новыми дополнительными свойствами, например, t – связные универсальные схемы.

Список литературы: 1. *Simmons G.J.* A game theory model of digital message authentication // Congressus Numerantium 34 (1992). С. 413 – 424. 2. *Simmons G.J.* Authentication theory/coding theory, in Advances in Cryptology // Proceedings of Crypto 84, Lecture Notes in Computer Science 196 (1985). С. 411 – 431. 3. *Wegman. M.N., Carter J.L.* New hash functions and their use in authentication and set equality // J. Computer and System Sci. 22 (1981).С. 265 – 279. 4. *Carter J. L., Wegman M. N.* Universal classes of hash functions // J. Computer and System Sci. 18 (1979). С. 143 – 154. 5. *Alon N., Goldreich O., Hastad J., Peralta R.* Simple constructions of almost k-wise independent random variables // Random Structures and Algorithms 3 (1992). С. 289 -- 304. 6. *Naor J., Naor M.* Small-bias probability spaces: efficient constructions and applications // SIAM Journal on Computing 22 (1993). С. 838 – 856. 7. *Helleseth T., Johansson T.* Universal hash functions from exponential sums over finite fields and Galois rings // Lecture Notes in Computer Science 1109 (1996). С. 31 – 44 (CRYPTO 96). 8. *Bierbrauer J., Schellwat H.* Weakly biased arrays, almost independent arrays and error-correcting codes // Publication in Proceedings of AMS-DIMACS (2000). 9. *Мак-Вильямс Ф.Дж., Слоэн Н.Дж.А.* Теория кодов, исправляющих ошибки. М.: Связь, 1979. 744 с. 10. *Халимов Г.З., Кузнецов А.А.* Аутентификация и универсальное хеширование // Радиотехника: Всеукр. межвед. науч.-техн. сб. 2001. Вып. 120. С. 100 – 110. 11. *Халимов Г.З., Кузнецов А.А.* Аутентификация с применением алгеброгеометрических кодов // Там же С. 103 – 109. 12. *Kurosawa K., Johansson T., Stinson D.* Almost k-wise independent sample spaces and their cryptologic applications // Lecture Notes in Computer Science 1233 (1997). С. 409 – 421. 13. *Carlitz L., Uchiyama S.* Bounds for exponential sums // Duke Mathematical Journal 24 (1957). С. 37 – 41.

Харьковский национальный
университет радиоэлектроники

Поступила в редколлегию 20.04.2003