# О сравнительном анализе двух схем формирования цифровых водяных знаков

Д. А. Мансурова*, В. В. Мкртичян
ФГАНУ НИИ "Спецвузавтоматика", Ростов-на-Дону, Россия
*Контактный e-mail: dianisia.m@yandex.ru

Многопользовательские схемы цифровых водяных знаков (ЦВЗ) предназначены для защиты авторских прав на цифровую продукцию. В данной работе рассматриваются две схемы ЦВЗ Боне — Шоу: базовая и каскадная, проводятся исследование и сравнительный анализ этих схем с целью определения более эффективной схемы из двух по длине, а также указаны границы значений параметров схем, для которых одна схема эффективнее другой.

*Ключевые слова*: защита авторских прав, цифровые водяные знаки, математические методы защиты информации.

## Введение

Ввиду развития информационных технологий и сети Интернет актуальным является вопрос защиты авторских прав файлов. Один из эффективных подходов защиты файлов от несанкционированного копирования и распространения — использование цифровых водяных знаков или цифровых отпечатков. Цифровой водяной знак (ЦВЗ) — определенная последовательность битов, добавляемая к цифровому документу, содержащая информацию об авторских правах создателя этого документа. Можно определить два вида схем ЦВЗ: однопользовательскую (watermarking) и многопользовательскую (fingerprinting). Основное отличие многопользовательской схемы заключается в уникальности внедряемого ЦВЗ для каждой копии тиражируемой продукции [1]. В [2] предложены две разновидности многопользовательской схемы ЦВЗ: базовая и каскадная. Цели настоящей работы — исследование двух этих схем, их сравнение и определение более эффективной схемы из двух предложенных с точки зрения длины метки, а также определение значений параметров схем, для которых это выполняется.

## Сравнительный анализ двух схем формирования цифровых водяных знаков

Пусть $\mathbb{N}$, $\mathbb{R}$ — множества натуральных и вещественных чисел соответственно; $F_2$ — поле Галуа; $l$, $n$, $c \in \mathbb{N}$; $F_2^l$ — линейное векторное пространство длины $l$ над полем $F_2$.

Множество $\Gamma = \{w^{(1)}, ..., w^{(n)}\} \subset F_2^l$ будем называть $(l, n)$-кодом, где $l$ — длина кода, $n$ — его мощность, $w^{(i)}$ — кодовое слово $(l, n)$-кода, являющееся ЦВЗ пользователя

с номером $i \in \{1; ...; n\}$. Рассмотрим некоторый $(l, n)$-код $\Gamma \subset F_2^l$. Коалицией кода $\Gamma$ будем называть такое подмножество $C$ этого кода, что выполняется условие $2 \leq |C| \leq n$.

Пользователем схемы ЦВЗ назовем легального владельца копии тиражируемой цифровой продукции. Предположим, коалиция $C$, состоящая из $c$ пользователей, создала копию продукта с нелегальным ЦВЗ [2]. Код, позволяющий определить хотя бы одного члена коалиции с вероятностью не менее $1-\varepsilon$, будем называть $c$-безопасным с ошибкой контролера $\varepsilon$, где $\varepsilon \in \mathbb{R}$.

Пусть $d$, $l$, $n \in \mathbb{N}$, $l = (n-1)d$. Базовым кодом $\Gamma_0(n, d) \subset F_2^l$ назовем код, состоящий из векторов $(w^{(1)}, w^{(2)}, ..., w^{(n)})$, где каждый $w^{(i)}$ представим в виде $w^{(i)} = (u_{i-1}, v_{n-i})$, где $i \in \{1; ...; n\}$; $u_j$ — нулевой вектор длины $jd$; $v_j$ — единичный вектор длины $jd$. Параметр $d$ будем называть размером блока кода.

Пусть $A$ — алфавит мощности $n$. Обозначим $A^L$ множество векторов длины $L$ над алфавитом $A$. Вспомогательным кодом $\Gamma'(L, N) \subset A^L$ назовем множество векторов $\{v^{(1)}; ...; v^{(N)}\}$ из $A^L$, где

$$v^{(i)} = (v_1^{(i)}, ..., v_L^{(i)}), \quad v_j^{(i)} \in A, \quad i \in \{1, ..., N\}, \quad j \in \{1, ..., L\}.$$

Пусть $\varphi$ — произвольное биективное отображение, $\varphi : A \longrightarrow \Gamma_0(n, d)$. Каскадным кодом $\Gamma_1(L, N, n, d)$ над кодами $\Gamma'(L, N)$ и $\Gamma_0(n, d)$ назовем множество векторов вида $(g^{(1)}, ..., g^{(N)})$, где $g^{(i)} = (g_1^{(i)}, ..., g_L^{(i)})$, и для любых $i$ и $j$, таких что $i \in \{1, ..., N\}$, $j \in \{1, ..., L\}$, выполняется условие $g_j^{(i)} = \varphi(v_j^{(i)})$, где $v_j^{(i)} \in A$.

В настоящей работе рассматриваются две схемы ЦВЗ: базовая и каскадная. Подробное описание этих схем приведено в [2]. В качестве внедряемых ЦВЗ для базовой схемы используются кодовые слова базового кода $\Gamma_0(n, d)$, а для каскадной — кодовые слова каскадного кода $\Gamma_1(L, N, n, d)$.

## Постановка задачи

Пусть $N$ — количество пользователей схемы, $\varepsilon$ — ошибка контролера, $c$ — размер коалиции. Рассмотрим коды $\Gamma_0(N, d_1)$ и $\Gamma_1(L, N, n, d_2)$ [2], где $d_1$ — размер блока базового кода, $d_2$ — размер блока базового кода в составе каскадного. Для этих кодов зафиксируем одинаковыми количество пользователей $N$ и ошибку контролера $\varepsilon$. Отметим, что согласно теоремам 12 и 17 из [2] схемы ЦВЗ с этими кодами будут функционировать корректно при следующих параметрах:

$$d_1 = 2N^2 \log(2N/\varepsilon), \quad d_2 = 2n^2 \log(4nL/\varepsilon), \quad n = 2c, \quad L = 2c \log(2N/\varepsilon).$$

Тогда длина кодов $\Gamma_0(N, d_1)$ и $\Gamma_1(L, N, n, d_2)$ вычисляется соответственно по формулам

$$l_1 = (N-1)d_1, \quad l_2 = (n-1)Ld_2.$$

Одну схему будем считать эффективнее другой по длине, если при одних и тех же входных параметрах $N$, $c$ и $\varepsilon$ длина кодовых слов этой схемы будет меньше. Сравним длину кодов $\Gamma_0(N, d_1)$ и $\Gamma_1(L, N, n, d_2)$ при одинаковом числе пользователей $N$ и ошибке контролера $\varepsilon$ для обеих схем. Определим, какая схема эффективнее по длине и при каких значениях параметра $c$ это выполняется.

**Формулировки полученных результатов**

**Утверждение 1.** Пусть заданы $N \in \mathbb{N}$ — количество пользователей, $c \geq 2 (\in \mathbb{N})$ — размер коалиции и $\varepsilon \in \mathbb{R}$ — ошибка контролера. Каскадная схема эффективнее базовой по длине, если выполняется неравенство

$$c(2c-1)d_2 < (N-1)N^2. \tag{1}$$

**Следствие.** Пусть заданы $c \in \mathbb{N}$, $N \in \mathbb{N}$ и $\varepsilon \in \mathbb{R}$. Положим $l_1 \in \mathbb{N}$ — длина кода $\Gamma_0(N, d_1)$, $l_2 \in \mathbb{N}$ — длина кода $\Gamma_1(L, N, n, d_2)$. Тогда каскадная схема эффективнее базовой по длине ($l_2 < l_1$), если выполняются следующие эквивалентные неравенства:

$$d_2/d_1 < (N-1)/((n-1)L),$$

$$c < \left\lfloor N^2(N-1)/d_2(n-1) \right\rfloor.$$

**Утверждение 2.** Пусть зафиксирована величина ошибки контролера $\varepsilon (\in \mathbb{R}) = 0.001$ и некоторые $c \in \mathbb{N}$, $N \in \mathbb{N}$. Если количество пользователей схемы $N$ меньше 15, то базовая схема эффективнее каскадной по длине.

Введем $b_1$ — бит эффективности каскадной схемы, который принимает значение 1, если каскадная схема эффективнее базовой по длине, и 0 в противном случае. Результат сравнения эффективности кодов $\Gamma_1$ и $\Gamma_0$ при $\varepsilon = 0.001$ в соответствии с утверждением приведен ниже.

Эффективность кодов $\Gamma_1$ и $\Gamma_0$ при $\varepsilon = 0.001$

| $N$ | $l_1$ | $c=1$ | | $c=2$ | | $c=3$ | |
|---|---|---|---|---|---|---|---|
| | | $l_2$ | $b_1$ | $l_2$ | $b_1$ | $l_2$ | $b_1$ |
| 5 | 1844 | 1824 | 1 | 47286 | 0 | 284760 | 0 |
| 7 | 5616 | 1920 | 1 | 50076 | 0 | 295510 | 0 |
| 9 | 12704 | 1920 | 1 | 51360 | 0 | 300900 | 0 |
| 13 | 41244 | 2037 | 1 | 52767 | 0 | 312015 | 0 |
| 15 | 64960 | 2037 | 1 | 54180 | 1 | 317440 | 0 |

**Утверждение 3.** Пусть зафиксировано количество пользователей $N \in \mathbb{N}$. Тогда максимальная мощность коалиции $c_{\max}$, при которой неравенство (1) справедливо, слабо убывает при быстром убывании $\varepsilon$. В частности, при $N = 300$ имеют место следующие значения: $c_{\max} = 18$ при $\varepsilon = 0.01$, $c_{\max} = 17$ при $\varepsilon = 0.001$, $c_{\max} = 17$ при $\varepsilon = 0.0001$, $c_{\max} = 16$ при $\varepsilon = 0.00001$.

## Заключение

В результате сравнительного анализа двух схем формирования ЦВЗ Боне — Шоу получены границы для параметров схем, при которых одна схема является эффективнее другой по длине (утверждения 1 и 2). В случае, когда число пользователей $N$ достаточно велико, а размер коалиции подразумевается небольшим, целесообразно использовать каскадную схему. Когда число пользователей $N$ небольшое или размер коалиции подразумевается достаточно большим, целесообразно использовать базовую схему. Также в результате численных экспериментов получено, что максимальная мощность коалиции, удовлетворяющая неравенству (1) при фиксированном количестве пользователей, зависит от величины ошибки контролера незначительно.

## Список литературы / References

[1] **Wagner, N.R.** Fingerprinting // Proc. of the 35th Annual ACM Symp. on Security and Privacy. 1983. P. 18–22.

[2] **Boneh, D., Shaw, J.** Collusion-secure fingerprinting for digital data // Lecture Notes in Comput. Sci. 1995. Vol. 963. P. 452–465.

*Поступила в редакцию 10 января 2018 г.,*
*с доработки — 28 апреля 2018 г.*

### On comparative analysis of two formation schemes of digital watermarks

Mansurova, Diana A.*, Mkrtichan, Vyacheslav V.

FSASI "Spetsvuzavtomatika", Rostov-on-Don, 344002, Russia

Corresponding author: Mansurova, Diana A., e-mail: dianisia.m@yandex.ru

Multiuser digital watermarking is a powerful tool for copyright protection of widely replicable digital products. This paper addresses Boneh — Shaw fingerprinting schemes. The algorithms proposed by D. Boneh and J. Shaw allow a distributor to detect any unauthorized copy and trace it back to user and they work correctly for any size of coalition. The main idea of this paper is a comparative analysis of two Boneh — Shaw digital watermarks schemes (cascade and basics), which makes clear which one of these schemes is more effective in terms of the length of the label. Using the comparative analysis and numerical experiments, we obtain the specific values of the boundaries for input parameters in which one scheme will be more efficient than another. From the results of the comparative analysis it can be concluded that for large number of users of the scheme and a small size of the coalition (in relation of users number) it is advisable to use the cascade scheme. In the case of a small number of users and a large size of the coalition (compared to the number of users) it is more expedient to use the basic scheme.

*Keywords*: copyright protection, digital watermarks, mathematical methods of information protection.

*Received 10 January 2018*
*Received in revised form 28 April 2018*