

ЛОКАЛЬНЫЙ И ГЛОБАЛЬНЫЙ ПОДХОДЫ К КОРРЕКТИРОВКЕ ВКРАПЛЕНИЙ В ДВОИЧНУЮ ПОСЛЕДОВАТЕЛЬНОСТЬ

В. А. ВОЛОШКО, Ю. С. ХАРИН, В. В. ПЬЯНОВ

Стеганография широко применяется для обеспечения защиты информации в сети Интернет. Главная цель стеганографии – незаметная передача сообщения. Одним из методов обнаружения факта передачи является статистический анализ контейнера, выявляющий аномалии. Поэтому после встраивания сообщения необходима маскирующая корректировка. В данной работе предлагаются два алгоритма корректировки – локальный и глобальный – для одномерного двоичного контейнера.

Алгоритм локальной корректировки. Локальная корректировка не меняет всю последовательность, а затрагивает лишь биты из некоторой окрестности мест вкрапления. Следуя [1], рассмотрим математическую модель вкраплений с однородной цепью Маркова первого порядка в качестве контейнера. Алгоритм локальной корректировки основывается на статистике [2]:

$$\Lambda(y) = \sum_{q \in \{0,1\}^2} \frac{(c_q - c_{q_1} p_q)^2}{c_{q_1} p_q}, \quad (1)$$

где c_q – вычисленные по наблюдаемой последовательности длины T частоты биграмм q ; p_q – вероятность перехода из q_1 в q_2 , $q = (q_1, q_2)$.

Теорема 1[2]. Критерий проверки гипотез асимптотического ($T \rightarrow \infty$) уровня значимости α о наличии вкраплений в однородную цепь Маркова первого порядка задается критической областью

$$\chi_{1\alpha} = \left\{ y \in \{0,1\}^T : \Lambda(y) \geq \chi_{1-\alpha,2}^2 \right\},$$

где $\chi_{1-\alpha,2}^2$ – квантиль уровня $1 - \alpha$ распределения χ^2 с двумя степенями свободы.

Определим k -окрестность символа x_i – вектор $(x_{i-k}, \dots, x_{i-1}, x_{i+1}, \dots, x_{i+k})$. Разработанный алгоритм локальной корректировки последовательно для каждой точки встраивания производит перебор всех 4^k возможных значений её k -окрестности и выбирает то значение окрестности, которое минимизирует статистику (1). Ширина окрестности выбирается так, что бы все окрестности всех точек встраивания не пересекались. Алгоритм производит корректировку до тех пор, пока не будет принята гипотеза об отсутствии вкраплений. Вычислительная сложность алгоритма имеет порядок $O(kT2^k)$.

В случае симметричной матрицы

$$\frac{1}{2} \begin{pmatrix} 1 + \varepsilon & 1 - \varepsilon \\ 1 - \varepsilon & 1 + \varepsilon \end{pmatrix} \quad (2)$$

переходных вероятностей контейнера можно рассмотреть статистику знакоперемен [1]:

$$B(y_1^T) = \sum_{t=1}^{T-1} y_t \oplus y_{t+1}, \quad (3)$$

где $\oplus$ – сумма по модулю 2.

Теорема 2[1]. Критерий проверки гипотез асимптотического уровня значимости α о наличии вкраплений в однородную цепь Маркова первого порядка с симметричной матрицей (2) переходных вероятностей задается критической областью

$$\chi_{1\alpha} = \{ y \in \{0,1\}^T : B(y) \geq B \},$$

где B – квантиль уровня $1 - \alpha$ биномиального распределения с числом испытаний $T - 1$ и вероятностью успеха $\frac{1}{2}(1 - \varepsilon)$.

В алгоритме локальной корректировки на основе статистики знакоперемен (3) можно избавиться от перебора и достичь вычислительной сложности порядка $O(kT)$.

Алгоритм глобальной корректировки. Глобальная корректировка означает полную замену битами специального процесса-корректора тех бит контейнера, ко-

торые не заняты передаваемым сообщением. Предполагается следующее расширение модели вкраплений [1]: процесс-селектор, выбирающий позиции для встраивания, и контейнер – стационарны и эргодичны. Корректор строится как псевдослучайная последовательность. Целью корректировки при этом является восстановление вероятностей и, как следствие, частот встречаемости всех двоичных слов некоторой наперед заданной длины n . Эти вероятности и частоты нужно сделать такими, какими они были в исходном контейнере до вкрапления. Опишем алгоритм построения корректора.

Шаг 1. Вычисление частот c_q , $q \in \{0,1\}^n$, слов длины n в контейнере.

Шаг 2. Построение по найденным частотам c_q и по вероятностям s_q слов длины n в селекторе так называемых характеристических функций (ХФ):

$$C_q = \sum_{q' \in \{0,1\}^n} c_{q'} (-1)^{H(qq')}, \quad S_q = \sum_{q' \in \{0,1\}^n} s_{q'} (-1)^{H(qq')}, \quad q \in \{0,1\}^n,$$

где $H(\cdot)$ – вес Хэмминга, число единиц в двоичном слове.

Шаг 3. Применение к ХФ селектора следующего преобразования:

$$S_q^* = 2^{-H(q)} \sum_{q' \leq q} S_{q'}, \quad q \in \{0,1\}^n,$$

где запись $q' \leq q$ означает соответствующее поэлементное неравенство.

Шаг 4. Вычисление ХФ корректора:

$$K_q = C_q / S_q^*, \quad q \in \{0,1\}^n.$$

Шаг 5. Вычисление вероятностей слов длины n корректора по его ХФ:

$$k_q = 2^{-n} \sum_{q' \in \{0,1\}^n} K_{q'} (-1)^{H(qq')}, \quad q \in \{0,1\}^n.$$

Шаг 6. Генерация корректора x_i в виде цепи Маркова порядка $n-1$ с распределением $\{k_q\}_{q \in \{0,1\}^n}$ слов длины n и переходными вероятностями:

$$P(x_t = Q | (x_{t-n+1}, x_{t-n+2}, \dots, x_{t-1}) = q) = \frac{k_{qQ}}{k_{q0} + k_{q1}}.$$

В самом распространенном случае, когда селектор является схемой независимых испытаний Бернулли с вероятностью $0 < \delta < 1$ выпадения единицы, шаги 2–5 алгоритма упрощаются, и вероятности $\{k_q\}_{q \in \{0,1\}^n}$ слов длины n корректора вычисляются по формулам:

$$k_q = \left(\frac{1-\delta/2}{1-\delta} \right)^n \sum_{q' \in \{0,1\}^n} c_{q'} \left(\frac{\delta}{\delta-2} \right)^{H(q \oplus q')}, \quad q \in \{0,1\}^n.$$

Корректор существует и решает поставленную задачу восстановления вероятностей, только если вычисленные величины $\{k_q\}_{q \in \{0,1\}^n}$ оказываются все строго положительными. В случае бернуллиевского селектора область допустимых значений доли вкраплений δ является интервалом $0 < \delta < \delta_*$, где граница δ_* зависит от контейнера. Наиболее «вместительным» контейнером с $\delta_*=1$ оказывается чисто случайная последовательность. Бернуллиевский контейнер с долей единиц d имеет границу $\delta_*=2\min(d, 1-d)$ для всех n . Как видим, эта граница уменьшается по мере отдаления процесса от чисто случайного. Та же тенденция наблюдается на практике при вычислении данной границы для реальных контейнеров, таких как последовательность младших бит пикселей изображений при LSB-встраивании. У «высокоэнтропийных» изображений значение δ_* оказывается большим, чем у структурированных.

Список использованных источников

1. Вечерко Е. В., Харин Ю. С. Статистическое оценивание параметров модели вкраплений в двоичную цепь Маркова. Дискрет. матем., 25(2): 135–148, 2013.
2. Billingsley P. Statistical Methods in Markov Chains. Ann. Math. Statist., 32(1): 12–40, 1961.